REMARKS

Claims 1-35 are cancelled. Claims 36, 38-40, 42, 43, 46-50, 52, 53, and 56-58 are amended. Claims 60-73 are withdrawn from consideration by the examiner.

Claims 36-57 and 59 stand rejected under 35 USC 103(a) as being unpatentable over Green et al (US 7,547,817) in view of Qui et al (Intl. Immunol. 1999; 11: 37-46).

Claim 58 stands rejected under 35 USC 103(c) as being unpatentable over Green et al in view of Qui et al and further in view of GenBank AC073553 (September 2002).

Claims 36 and 52 stand rejected under 35 USC 103(a) as being unpatentable over Green et al in view of Qui et al and further in view of Harriman et al.

Each of these rejections are addressed and traversed in the following remarks.

Background of the of the invention

The IgH locus structure and the mechanism of class switch recombination are well-known in the art (see *Immunology, Ivan Roitt et al., Third Edition, 1993, Mosby, pages 5.9 to 5.11 and 11.9 to 11.12; Annex I*). B cell maturation requires the expression of membrane IgM on B cells surface. The production of a selected antibody isotype requires the development of IgM producing B cells which subsequently undergo isotype class switching from IgM to the IgG or IgA, or IgE isoptypes. Isotype class switch is mediated through a deletional recombination event (class switch recombination or CSR) occurring between tandem directly repetitive switch regions (S) present 5' of all IgH constant region genes except Cδ. Transcription starting from an I exon located immediately upstream of the switch region and continuing through the switch region is required for CSR. Enhancers and cytokine response sequences are known to lie in the region near the I promoter.

Brief description of the claimed invention

The claimed invention relates to a transgenic knock-in mammal, wherein the endogenous immunoglobulin heavy chain locus (IgH) has been modified by homologous recombination to replace the switch sequence Sμ with a transgene containing a human heavy chain constant region gene Cα or a segment of the Cα gene containing at least an exon encoding the CH3 domain and a membrane exon.

Figure 1 of the present specification illustrates the endogenous IgH locus in mice, the targeting vector which can be used for homologous recombination and the murine locus after homologous recombination.

The deletion of the switch sequence Sμ in the endogenous IgH locus and its replacement with the human Cα transgene produces the following effects on the immunoglogulin heavy chain genes expression:

(i) a chimeric human IgA heavy chain in which the constant region is human and the variable region is from the non-human mammal is expressed to high level; this chimeric IgA heavy chain benefits from a completely diversified repertoire corresponding to the normal repertoire generated by rearrangements of the VH, D and JH segments at the non-human mammal IgH locus,

(ii) the expression of the endogenous Cμ gene responsible for the synthesis of IgM heavy chains is abolished, and

(iii) the expression of the other immunoglobulin heavy chains genes which code for the other classes of immunoglobulins (IgG and IgE) is greatly reduced due to the blocking of the Ig class switch toward the immunoglobulin constant genes located downstream of Cμ on the endogenous IgH locus.

These specific characteristics of the non-human transgenic mammal of the claimed invention are demonstrated in the examples. For example, Example 1 demonstrates that the non-human transgenic mammal of the claimed invention produces antibodies which contain large quantities of chimeric human IgAs (in the gram per litre range in mice) but no IgM. Example 3 demonstrates that the transgenic mammal is capable of producing

antibodies with high affinity as a secondary response to the antigen since its B lymphocytes can recruit the somatic hypermutation phenomenon.

The brief description of the claimed invention above is intended solely for purposes of clarification and is not intended to limit either the scope of any pending claims.

Green *et al.* (US 7,547,817)

Green describes human antibodies producing transgenic mammals (Xenomouse™) which are obtained by introducing yeast artificial chromosome cloning vectors containing large germline fragments (unrearranged) of the human Ig locus (human Ig YAC transgene), into the somatic and germline cells of a mammal (column 2, line 31 to column 3, line 8 and column 4, lines 4-14).

Green describes that current technologies for obtaining a transgenic mouse which produces an antibody of the desired isotype requires antibody re-engineering *in vitro* which is labor intensive, slow and expensive (column 4, line 60 to column 5, line 36).

Green is thus directed to solving the specific problem of obtaining a pre-selected human antibody isotype from a transgenic mouse (column 5, lines 53-55).

Green describes that due to the differential responsiveness of mouse and human switch regions to lymphokines and other activators it is desirable to have heterologous switch regions controlling CSR in human antibody producing mice (column 12, lines 29-33).

Green discloses a transgenic mammal containing in its somatic and germline cells an unrearranged human immunoglobulin heavy chain (IgH) YAC transgene containing VH genes, all the D elements, all J elements, Sμ, Cμ and Cδ from human chromosome 14, and a chimeric human IgH constant region gene in which human constant region gene sequences encoding the desired heavy chain isotype are operably linked to an

heterologous or non-cognate switch region (column 5, line 66 to column 6, line 3; column 6, lines 21 to 36). For example, a mouse switch region is operably linked to a human gamma, alpha or epsilon constant region coding segment, or a human switch region is operably linked to a human constant region segment, the switch region being from a different isotype than the constant region coding segment (column 6, lines 40-45 and 51-55).The heterologous switch region controls switching from the human IgM to the downstream human Ig of the desired isotype. The transgenic mammal of Green is engineered to produce human IgM and human Ig of the desired isotype (column 5, line 66 to column 6, line 3).

A first transgenic mouse line producing human IgM and IgG2 antibodies (Xenomouse) was previously derived by Green, from the yH1C transgene composed of 66 VH genes, all the D elements, all J elements, Cμ and Cδ and all regulatory elements in germline configuration, appended in 3' with a 22 kb fragment containing the human Cγ2 gene, including its switch region (Sγ2), and a 4 kb fragment containing the mouse 3' enhancer element (column 9, line 58 to column 10, line 42).

The human IgH YAC transgenes disclosed in Green are engineered by introducing a targeting vector having 5' and 3' flanking homology to yH1C and an appropriate selection marker, into yeast carrying yH1C. Such vector can be recombined *in vivo* in yeast to replace the human Sγ2 and Cγ2 with the mouse Sγ1 functionally linked to a human CH coding sequence (*e.g.* human or the human Cγ1) or the human Cγ2 by another human CH gene, in yeast carrying yH1C (column 12, line 36 to column 13, line 9 and figures 1 to 7). To produce an antibody of the desired isotype, new transgenic Xenomouse are then generated by introducing the recombinants IgH YAC transgenes into mouse ES cells.

Harriman *et al.*, J. Clin. Invest., 1996, 97, 477-485 and Qiu *et al.*, Int. Immunol., 1999, 11, 37-46

These two research papers from Harriman are directed to understanding the process of antibody class switching and their implication in IgA deficiency (abstract and introduction of Harriman *et al.*). For this purpose, an Iα knock-out mouse was generated by targeted deletion (Figure 1 of Harriman *et al.*) and used in Qiu *et al.* (figure 1 and page 38, 2nd column, beginning of second paragraph which refers to Harriman). In this Iα knock-out mouse, a human HPRT mini-gene driven by the PGK promoter and having a SV40 poly(A) signal replaces the Iα exon and proximal promoter located upstream of the switch region Sα, in the endogenous IgH constant region gene Cα. The knock-out mouse is generated by targeted deletion. Therefore, all the sequences 5' and 3' to the Iα exon which include Sμ and Cμ in 5' and Sα in 3', are present in the knock-out mouse IgH locus. Harriman *et al.* teaches that the Iα knock-out mouse produces mouse IgM, IgG and IgA at levels equivalent to those of the wild-type (page 480, 2nd column, 2nd paragraph and page 481, 1st column, end of first paragraph). Harriman *et al.* teaches also that the Iα exon or transcripts containing the Iα exon are not required for IgA class switch but a second signal is required for the induction of IgA class switch (abstract). Qui *et al.* discloses that a transcript of any sequence which is spliced across the switch region is necessary and maybe sufficient for CSR.

GenBank AC073553.5

GenBank AC073553.5 discloses the sequence of a 187523 bp DNA segment from mouse chromosome 12.

Non-obviousness of claims 36-57, 59 over Green et al. in view of Qiu et al.

The rejection of claims 36-57 and 59 under 35 USC 103(a) over Green *et al.* (US 7,547,817) in view of Qiu *et al.* (Int. Immunol., 1999, 11, 37-46) is respectfully traversed as the cited references do not disclose all the elements of the claimed invention, suggest the combination of elements of the claimed invention, or provide a reasonable expectation of success for the claimed invention.

Green was cited as differing from the present claims in that although Green may provide transgenic mice for producing any desired isotypes of human antibodies including IgA, the exemplified target IgH gene is not Cα (page 6 of the present Office Action).

However this is respectfully untrue. It is worth noting that the Examiner herself acknowledges that it is not true since it is noted on page 8 of the present Office Action that "*the combined teaching of Green in view of Qiu teaches a transgenic mouse with a non-cognative S region in its genome but not lacking a Sμ*".

Green differs from the present claims in that Green, at the very least, does not disclose or suggest any of the following features from the claimed invention:

1. Green does not disclose a knock-in mammal generated by homologous recombination at the mammal endogenous IgH locus but a transgenic mammal (Xenomouse™) generated by random insertion of large germline fragments of the human IgH locus, into the somatic and germline cells of a mouse using YAC vectors (Column 5, Line 66 to Column 6, Line 3, Column 6, Lines 36-37;Column 7, lines 44-48; Column 9, Line 58 to Column 10, Line 42).
   Therefore, contrary to what is stated by the Examiner on page 6 of the present Office Action, Green does not disclose a targeting vector comprising 5' and 3' mouse flanking sequences for homologous recombination, which is introduced into mouse ES cells.

2. Green does not disclose targeted DNA sequence replacement at the mouse endogenous IgH locus. Green discloses targeted DNA sequence replacement in yeast, on a human IgH YAC transgene comprising unrearranged variable region genes (V, D, J genes) Sμ, Cμ, Cδ, Sγ2, Cγ2, and a 4 kb fragment containing the mouse 3' enhancer element (column 12, line 36 to column 13, line 9, cited by the Examiner on page 5 of the present Office action, and figures 1 to 7).

3. Green does not disclose targeted DNA sequence replacement, wherein the endogenous switch sequence Sμ is replaced with a transgene construct comprising a human heavy chain constant region gene Cα or a segment of the Cα gene comprising at least an exon encoding the CH3 domain and a membrane exon.

   Green discloses targeted DNA sequence replacement, wherein: (i) the human Sγ2 switch sequence and the human Cγ2 coding sequence are replaced with the mouse Sγ1 switch sequence functionally linked to a human CH coding sequence (*e.g.* human or the human Cγ1), or (ii) the human Cγ2 coding sequence is replaced by another human CH gene (column 6, lines 40-45 and 51-55; column 12, line 36 to column 13, line 9 (cited by the Examiner on page 5 of the present Office action) and figures 1 to 7).

   Therefore, contrary to what is stated by the Examiner on page 5 of the Office action, Green does not disclose the deletion of the mouse Sμ since Green discloses targeted replacement within a human immunoglobulin heavy chain (IgH) transgene. In addition, the targeted replacement disclosed by Green does not contain the replacement of Sμ but of Cγ2 alone or together with its switch region Sγ2.

4. Green does not disclose a human heavy chain constant region (CH) transgene which is not linked to a switch region. Green discloses a human CH gene which is always functionally linked to an heterologous switch region (column 6, Lines 29 to 32 and column 12 cited by the Examiner on page 5 of the present Office Action).

5. Green does not disclose the insertion of a human CH transgene between the intronic enhancer Eμ and the Cμ gene (figure 1 of the present application) but downstream of the Cμ gene (figures 1 to 4 of Green).

6. Green does not disclose a modified IgH locus that is incapable of isotype switching from IgM to the isotype of the downstream human CH transgene. Green discloses a human IgH transgene which is always functionally linked to an heterologous switch

region and capable of isotype switching from the human IgM to the isotype of the downstream human CH gene (column 6, lines 26 to 29 and column 12 cited by the Examiner on page 5 of the present Office Action).

7. Green does not disclose a transgenic mammal which produces no IgM and high level of chimeric human IgAs in which the heavy chains comprise a human immunoglobulin A constant region and a variable region from the non-human mammal. Green discloses a transgenic mammal which produces human IgM and human Ig of the desired isotype.

Qiu was cited as disclosing, on figure 1, a transgenic mouse whose endogenous switch region and Cα region was replaced with a human Sα and Cα (page 6 of the present Office Action). However this is respectfully untrue. As mentioned above in paragraph 3.2.3, Qiu discloses an Iα knock-out mouse. In this mouse, the endogenous switch region and Cα region are not modified and there is no human Sα and Cα insertion (figure 1 of Qiu or Harriman which both refer to the same mouse).

One of ordinary skill in the art would have no reason to combine Green with Qiu because Qiu does not disclose or suggest any element of the claimed invention which is missing from Green. Qiu discloses an Iα knock-out mouse that produces mouse IgM, IgG and IgA at levels comparable to the wild-type, as discussed above.

Even, if the ordinary skilled artisan were to combine the disclosure of Green with that of Qiu, the ordinary artisan would modify the mouse described by Green by linking the heterologous switch region functionally to an HPRT minigene as taught by Qiu to arrive at a transgenic mammal containing an unrearranged human immunoglobulin heavy chain (IgH) transgene containing VH genes, all the D elements, all J elements, Sμ, Cμ and Cδ from human chromosome 14, and a chimeric human IgH constant region gene in which human constant region gene sequences encoding the desired heavy chain isotype are operably linked to an heterologous switch region which is operably linked to an

HPRT minigene, *i.e.* a transgenic mammal which is even more different from the claimed transgenic mammal than the transgenic mammal of Green.

Furthermore, the prior art and the common knowledge at the time of filing of the present application (*Annex I*) , in fact, teach away from the present invention. It is known in the art that B cell maturation, *in vivo*, requires the expression of membrane IgM on B cells surface and that the production of a selected antibody isotype, *in vivo*, requires the development of IgM producing B cells which subsequently undergo isotype class switching from IgM to the IgG or IgA, or IgE isoptypes (column 10, lines 43 to 60 of Green and *Annex I*). As discussed above, the mice which are disclosed in Green, Qiu and Harriman (which refer to the same knock-out mouse as Qiu) have IgM producing B cells which subsequently undergo isotype class switching from IgM to the IgG or IgA, or IgE isoptype. Therefore, to produce human Ig of the desired isotype (IgG, IgA,) one of ordinary skill in the art would never have engineered mice which do not produce IgM but only Ig of another isotype (IgG, IgA). In accordance with the claimed invention, the present inventors have demonstrated that, in the absence of IgM expression, mammals can surprisingly develop functional B cells able to generate an immune response.

Therefore, even the combined disclosures of the cited references would fail to disclose all the elements of the claimed invention, or suggest this combination or provide a reasonable expectation of success for the present invention because numerous elements from the claimed invention are missing from these reference disclosures. In fact, these references actually teach away from the present invention.

Hence, this ground of rejection is unsustainable and should be withdrawn.

Non-obviousness of claim 58 over Green et al. in view of Qiu et al. and GenBank AC073553.5

Claim 58 stands rejected under 35 USC 103(a) as being unpatentable over Green *et al.* (US 7,547,817), in view of Qiu *et al.* (Int. Immunol., 1999, 11, 37-46) as applied to claims 36-57, 59, and further in view of GenBank AC073553.5.

This rejection is not sustainable over the combination of Green and Qiu for the reasons discussed above. However, GenBankAC073553.5 does not remedy the deficiencies of Green and Qui.

GenBank AC073553.5 discloses the sequence of a 187523 bp DNA segment from mouse chromosome 12.

However, GenBank AC073553.5 does not disclose the specific fragments corresponding to SEQ ID NO: 7 and SEQ ID NO: 8 of the invention, nor the combination of said specific fragments with a human Cα gene or a fragment of said gene comprising the CH3 domain and membrane exons. Furthermore, there is no suggestion of such a combination, nor does this reference provide a reasonable expectation of success the same because the conventional knowledge at the time of filing of the present application actually teaches away from the claimed invention for the reasons discussed above.

Accordingly, this rejection is unsustainable and should be withdrawn.

Non-obviousness of claims 36 and 52 over Green et al. in view of Qiu et al. and Harriman et al.

Claims 36 and 52 stand rejected under 35 USC 103(a) as being unpatentable over Green *et al.* (US 7,547,817), in view of Qiu *et al.* (Int. Immunol., 1999, 11, 37-46) as applied to claims 36-57, 59, and further in view of Harriman *et al.* (J. Clin. Invest., 1996, 97, 477-485).

Harriman was cited as establishing that it was known in the art that the Sμ region is not required for IgA class switch, and that a mouse without the genomic Sμ region still can produce IgA antibody.

However this is not true for the reasons discussed above. Harriman and Qiu disclose the same Iα knock-out mouse. In this knock-out mouse, generated by targeted deletion, the only region of the endogenous IgH locus which is deleted is the Iα exon (figure 1 of Harriman). Therefore, all the sequences 5' and 3' to the Iα exon which include Sμ and Sγ in 5' and Sα in 3' are present in the Iα knock-out mouse IgH locus. Harriman teaches that the Iα knock-out mouse produces IgM, IgG and IgA and that the IgA are produced by class switching, *i.e.*, by recombination between the Sμ and Sα, or Sγ and Sα regions.

Therefore, Harriman does not disclose or suggest that the Sμ region is not required for IgA class switch, and that a mouse without the genomic Sμ region still can produce IgA antibody.

Importantly, Harriman fails to disclose all of the elements missing from the two primary references, or suggest this combination or provide any reasonable expectation of success for the present invention. In fact, the combined disclosures of the cited references would teach away from the claimed invention for the reasons noted above.

Hence, this ground of rejection is unsustainable and should be withdrawn.

Claims 36-51 and 59 stand rejected under 35 USC 112, first paragraph.

However, the rejection of claims 36-51 and 59 under 35 USC 112, first paragraph, as failing to comply with the written description requirement is respectfully traversed as the IgH locus sequences required for practicing the claimed non-human mammal are described in the specification in such a way as to reasonably convey to one skilled in the

relevant art that the inventors, at the time the application was filed, had possession of the claimed invention.

The practicing of the claimed invention does not require all the sequences of the IgH locus but only a very small portion of sequences (less than 10 kb) flanking the Sμ region, the targeted site for homologous recombination (see figure 2).

These sequences which are necessary and sufficient for practicing the claimed invention are defined by specific features (size, well-known protein domains, *i.e.*, JH/Eμ region, Cμ gene) which are disclosed in the specification (see figure 2).

The specification provides also the positions of these sequences on mouse chromosome 12 sequence (Genbank/EMBL AC073553).

At the date of filing of the present application, the IgH gene locus had been mapped in different mammal species and its entire sequence or part of its sequence comprising the JH/Eμ region and Cμ gene sequences were available for numerous species, including at least human, mouse, rat, sheep, cattle, dog, cat, rabbit, hamster, shrew and pig (Schrenzel *et al.*, Immunogenetics, 1997, 45, page 386, 2nd column, end of last paragraph; *Annex II*). Examples of GenBank accession numbers are: AY158087 (bovine), M13800.1 (rat) and X02804.1 (hamster), shown in *Annexes III to V*.

Furthermore, using the mouse sequences that are described in the application, one skilled in the art would have been able to obtain the corresponding sequences from other mammals, using standard molecular biology and/or sequence analysis techniques which were routine in the art at the date of filing of the present application.

Therefore, this ground of rejection is clearly moot.

Claims 36-51 and 59 stand rejected under 35 USC 112, first paragraph.

However, the rejection of claims 36-51 and 59 under 35 USC 112, first paragraph, as failing to comply with the enablement requirement is respectfully traversed as the specification does enable any person skilled in the art to which it pertains, or with it is most nearly connected, to practice the invention commensurate in scope with these claims.

First of all the essential materials (IgH sequences) required for practicing the claimed invention are adequately described by the instant disclosure for the reasons mentioned above.

The specification describes how the claimed transgenic mammal can be obtained by homologous recombination with an appropriate targeting vector (see page 11, lines 4-11, page 11, line 31 to page 12, line 14) and provides a working example of the invention (transgenic mouse line of example 1).

The state of the art and the level of skill in the art is such that pronuclear microinjection of fertilized eggs and the use of *in vitro* embryo production in combination with gene transfer technology are available for the mammal species for which ES cells were not available (Mullins *et al.*, J. Clin. Investigation, 1996, 97, 1557-1560; in particular page 1557, second column, end of paragraph entitled "*Transgenesis by pronuclear injection*").

The inefficiency of pronuclear microinjection due to random integration of the transgene is not a relevant problem for the present invention since the transgene is integrated by homologous recombination in a functional locus. Mullins *et al.* mentions clearly that targeted homologous recombination improves the inefficiency of pronuclear microinjection (see page 1558, first column, middle of first paragraph: "*In many cases, simply including large amounts of flanking sequences may be sufficient to overcome position effects and direct expression to specific tissues*"). Therefore, pronuclear microinjection is not an unpredictable technique in the case of the present invention

since it allows efficient integration of the transgene at the mammal IgH locus and subsequent expression of this transgene.

Furthermore, Mullins *et al.* point out that the use of *in vitro* embryo production in combination with gene transfer technology is a major improvement to pronuclear microinjection of fertilized eggs because transgene screening and cloning can take place before reintroduction into the natural host. Mullins et al. mentions that the microinjected embryo technology allows efficient production of transgenic cattle, rabbit and sheep.

Therefore, one skilled in the art with the teaching of the specification and the knowledge in the art would be able to practice the claimed transgenic mammal without undue experimentation.

Therefore, this ground of rejection is deemed to be moot.

Claims 36-51 stand rejected under 35 USC 112, second paragraph.

However, the rejection of claims 36-51 are believed to be obviated by the amendments which were made to claims 36, 46 and 48.

Therefore, withdrawal of the rejection is respectfully requested.

The specification stands objected to.
In view of above amendment to the specification, this ground of rejection is moot.

Finally, Applicants emphasize that even assuming for the sake of argument that one skilled in the art would have had motivation to combine the cited references of record, the artisan still would not have been put in possession of the claimed invention. Moreover, KSR v. Teleflex, 550 U.S. 398 (2007), is of no moment to the patentability of the claimed invention since in the present case one skilled in the art could not have

relied upon "common sense" to find implicit motivation to combine the cited references for at least three reasons.

First, even the combined reference disclosures lack the full measure of claimed genetic elements. That is, the combined teachings fall short of the claimed invention.

Second, the combined reference disclosures actually teach away from the claimed invention. Unlike in KSR, where the U.S. Supreme Court found nothing in the prior art that would teach away from the Asano reference (used to determine the obviousness of the claimed invention), on the present record there is cited prior art that teaches away from the claimed invention.

Third, the claimed invention does not constitute a known problem for which there is a known solution. In fact, to the contrary, the cited references strongly indicate that the presently claimed 'solution' was anything but known, and, thus patentable.

Guidelines for U.S. Patent Examiners are set forth at MPEP 2143, Eighth Edition (Revision 6, September, 2007). Notably, "Exemplary Rationales" A-G are set forth where an examiner might justifiably cite the KSR decision in support of an obviousness rejection. In fact, none of these rationales apply to the claims at issue. This is striking inasmuch as the KSR test for determining obviousness is clearly broader than the previously (and solely) used 'teach, suggest, motivate' (TSM) test. Yet, even an attempted application of the KSR test indicates patentability for the claimed invention based upon the prior art of record.

Annexes I-V are attached to this Amendment in general support of the patentability of the claimed invention. As such, they have not been cited in an Information Disclosure Statement in accordance with standard practice.

Accordingly, in view of all of the above amendments and related remarks, it is believed this application is now in condition for allowance. Favorable consideration and early notice to this effect are earnestly solicited.

Respectfully submitted,

William E. Beaumont
Reg. No. 30,996
Juneau Partners, PLLC
Customer No.: 50438

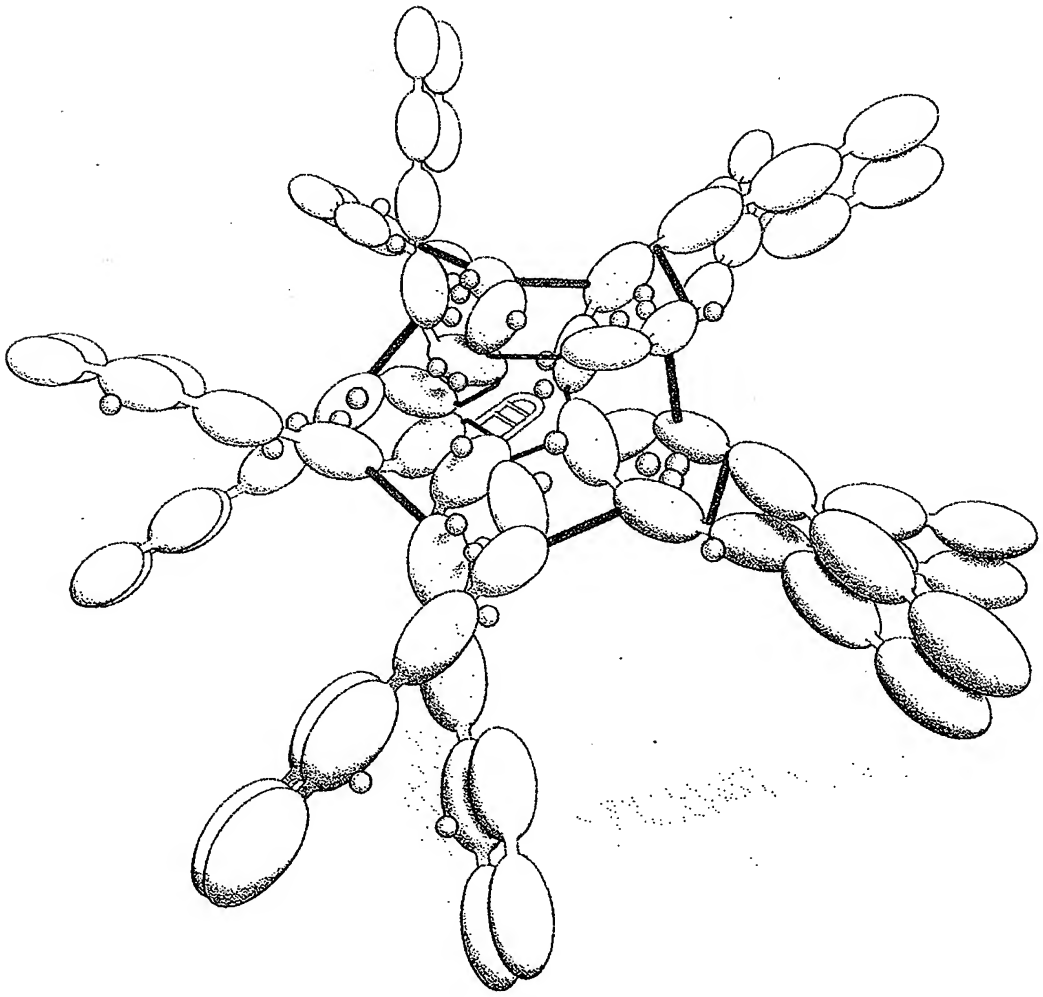
* See Annexes I-V attached to this amendment

Annex I

IMMUNOLOGY

THIRD EDITION

Ivan Roitt • Jonathan Brostoff • David Male



Birds

In contrast to sharks, chickens possess extremely limited numbers of genes coding for immunoglobulins. In the light chain system there is only one V, one J and one C segment gene. The heavy chains are similarly restricted with single V and J segments. Although there are about 15 DH segments, these are all very similar in sequence and add little to diversity. Despite this severe limitation chickens are perfectly able to mount a wide range of antibody responses and produce sequentially diverse antibody molecules.

Upstream of the VL gene is a region containing 25 sequences similar to VL regions, but each lacking a leader exon and a promoter region. They also lack the characteristic heptamer–spacer–nonamer sequences needed for V–J rearrangement. These pseudogenes are not wasted but are used in a process of gene conversion, with sections of the pseudogene being inserted into the viable VL region. This is a continuous process which carries on after the B cells have left the bursa and multiple conversion events can occur during the lifetime of the B cell.

Mutations in the DNA of two VH T15 genes

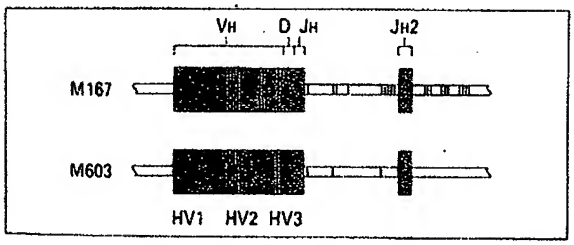


Fig. 5.16 DNA of two anti-phosphorylcholine antibodies with the T15 idiotype. (Black lines indicate positions where the genome has mutated from the germ line sequence.) There are large numbers of mutations in the introns and the exons of both genes, but particularly in the second hypervariable region, HV2. By comparison, no mutations are detectable in genes coding for the constant regions.

Five mechanisms for the generation of antibody diversity

1. multiple germ line V genes
2. V–J and V–D–J recombinations
3. recombinational inaccuracies
4. somatic point mutation
5. assorted heavy and light chains

Fig. 5.17 Since each mechanism can occur with any of the others, the potential for increased diversity multiplies at each step of immunoglobulin production.

Similar processes occur with the heavy-chain gene locus, where up to a 100 VH pseudogenes act to increase the diversity by similar conversion mechanisms.

Rabbits

Rabbit immunoglobulins have always presented a puzzle, particularly in the way in which allotypes are regulated. Although the rabbit has many VH genes, the VH gene nearest to the D segment is used in most rabbit B cells. Recent evidence suggests that the rabbit may also use a gene conversion mechanism to diversify this single VH gene.

PSEUDOGENES IN HUMAN DIVERSIFICATION

Several V and J segment genes are also in the form of pseudogenes. Whether gene conversion is involved in generating human V regions is a matter of speculation and interest.

HEAVY CHAIN CONSTANT REGION GENES

All classes of immunoglobulin use the same set of variable region genes. When the class is changed, by a B cell that has matured into an antibody-forming cell and is therefore committed to a particular antigen, all that is switched is the constant region of the heavy chain. This is also shown by the analysis of double myelomas, where two monoclonal antibodies are present in the serum at the same time. IgM and IgG antibodies from a patient with multiple myeloma have been found to have identical light chains and VH regions; only the constant regions were switched from μ to γ. Often IgM and IgD are found on the lymphocyte surface at the same time. Capping these receptors with antigen has revealed that the IgM and IgD from the same B cell have the same specificity for antigen, indicating similarity of VH regions on the two classes (Fig. 5.18).

All the constant region genes are arranged downstream from the J segment genes. In the mouse there is one gene for each of the μ, ϵ and α isotypes and one γ gene for each of the four different IgG isotypes (Fig. 5.19). In man the constant region genes are more complicated, and it appears that one section of this region has undergone gene duplication and diversification. In man the ordering of genes is μ, δ, [γ3, γ1, ϵ1 α1], γ, [γ2, γ4, ϵ, α2]. The two sets within square brackets indicate the possible area of reduplication. The genes ϵ1 and γ are pseudogenes and are not expressed. Just upstream (5') to the μ genes is a switch sequence (S) which is repeated upstream (5') to each of the other constant region genes (Fig. 5.20). Class switching is important in the maturation of the immune response and may be accompanied or preceded by somatic mutation. Initially a complete section of DNA, including the recombined VH region through the δ and μ constant regions, is transcribed; then by differential splicing, two mRNA molecules are produced, each with the same VH but having either μ or δ constant regions. It is suggested that sometimes much larger stretches of DNA are also

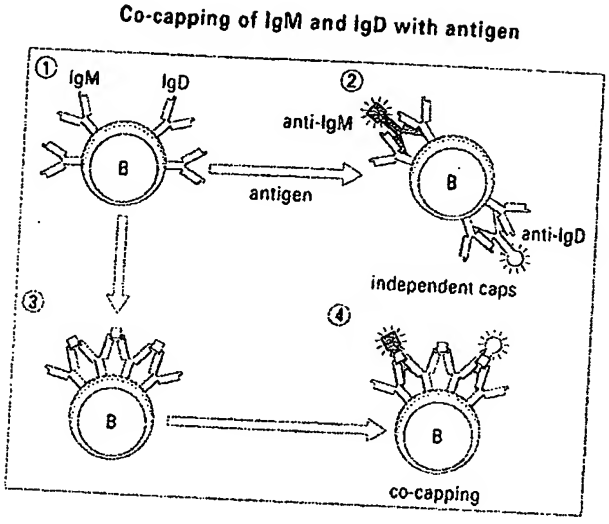


Fig. 5.18 Some B cells have both IgM and IgD on their surface (1). This can be demonstrated by treating the cells with rhodaminated anti-IgM (red) and fluoresceinated anti-IgD (green). The anti-IgM causes the IgM antibodies to aggregate on the surface, while the anti-IgD causes the IgD to aggregate, producing separate red and green caps on the cell (2). If the experiment is repeated by first treating the cells with antigen (blue) as in (3) and then with the anti-IgM and anti-IgD, the red anti-IgM and green anti-IgD caps appear together on the cell, that is, they co-cap. This implies that the IgM and IgD on the cell surfaces are cross-linked by antigen (4). This could only occur if the IgM and IgD had the same antigen-binding specificity, and is therefore evidence that different constant regions (μ and δ) can be linked to the same V region.

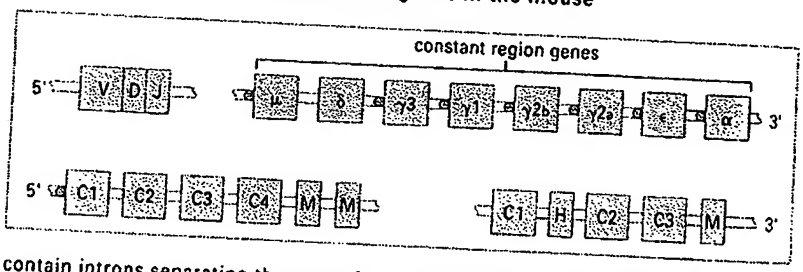


Fig. 5.19 The constant region genes of the mouse are arranged 6.5 kb downstream from the the recombined V-D-J segment. Each C gene (except Cδ) has one or more switching sequences at its start (red circles) which correspond to a sequence at the 5' end of the μ gene. This allows any of the C genes to recombine with V-D-J. δ genes appear to use the same switching sequences as μ but the μ gene transcript is lost in RNA processing to produce IgD. The C genes (expanded below for μ and γ2a) contain introns separating the exons for each domain (C1, C2, etc.). The γ genes also have a separate exon coding for the hinge (H), and all the genes have one or more exons coding for membrane-bound immunoglobulin (M).

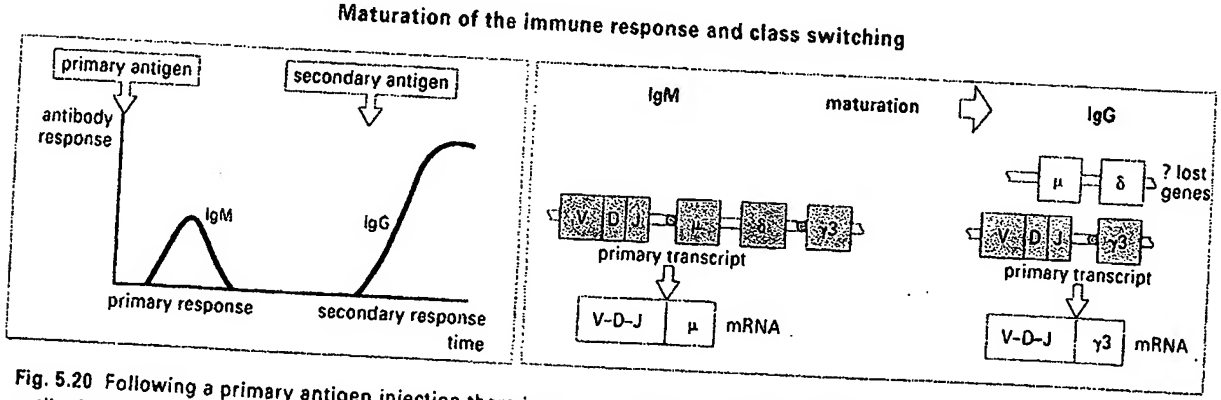


Fig. 5.20 Following a primary antigen injection there is an antibody response which consists mostly of IgM; the response following a secondary challenge is mostly IgG. The underlying mechanism for the class switch is shown (right). In the primary response the V-D-J region is transcribed with a μ gene. After removal of introns during processing, mRNA for secreted IgM is produced. During maturation (involving T-cell help and possibly activation of a mutation mechanism for the V-D-J segment) another C gene ($C_{\gamma 3}$ in this case) is brought up to exchange with the μ gene at its switch region (red). The μ and δ genes are probably lost; transcription and processing produce mRNA for IgG3.

Isotype switching by differential RNA splicing

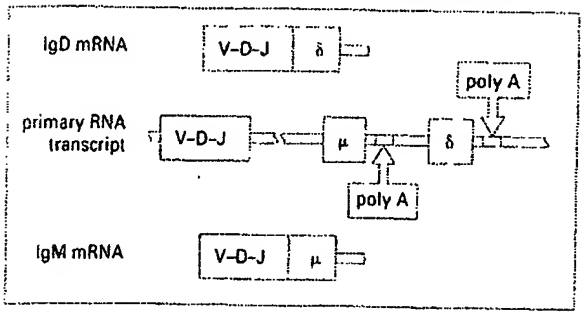


Fig. 5.21 Single B cells produce more than one antibody isotype from a single long primary RNA transcript. A transcript containing μ and δ is shown here. Polyadenylation can occur at different sites (black), leading to different forms of splicing, producing mRNA for IgD (top) or IgM (bottom). Even within this region there are additional polyadenylation sites which determine whether membrane immunoglobulin or secreted immunoglobulin is formed.

Lost genes

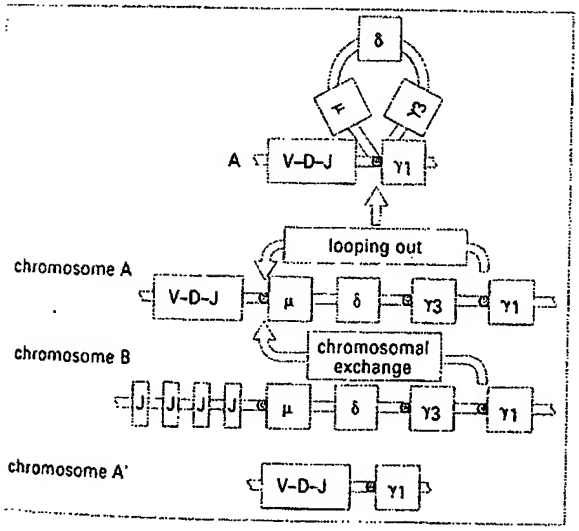


Fig. 5.22 A and B are chromatids of the chromosome section of the Ig genes. A contains the rearranged V-D-J segment. In the looping out hypothesis, a section of C genes (μ, δ, γ3) loop out and are lost. In the chromosome exchange hypothesis, similarities in the switching sequences permit unequal somatic recombination between maternal and paternal chromosomes. The A chromosome recombines with another part of the unrearranged B chromosome. The loss of gene segments gives rise to the IgM-IgG1 switch, shown as A'. The 'lost' C genes are on the other, non-functional chromosome B' (not shown), which now contains two copies of several C genes.

transcribed together, with differential splicing giving other immunoglobulin classes sharing V_H regions (Fig. 5.21). This has been observed in cells simultaneously producing IgM and IgE. More often, class switching appears to be mediated by a recombination between S recombination sites, allowing a looping out and deletion of DNA and bringing another C region close to the VDJ gene (Fig. 5.22). A further possibility has been suggested involving exchange between chromosomes.

MEMBRANE AND SECRETED IMMUNOGLOBULIN

Membrane immunoglobulin (antigen receptor) is identical to secreted immunoglobulin (antibody), except for a stretch of amino acids at the C terminus of each heavy chain. Membrane immunoglobulins are larger than their

Membrane and secreted IgM: mouse

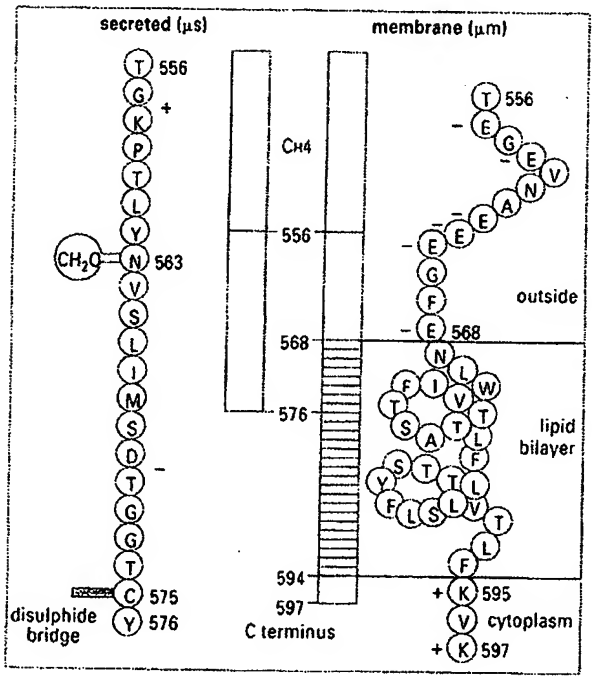


Fig 5.23 C-terminal amino acid sequences are shown for both secreted and membrane-bound IgM, and are identical up to residue 556. Secreted IgM has 20 further residues. Residue 563 (asparagine) has a carbohydrate unit attached to it while residue 575 is a cysteine involved in the formation of interchain disulphide bonds. Membrane IgM has 41 residues beyond 556. A stretch of 26 residues between 568 and 595 contains hydrophobic amino acids sandwiched between sequences containing charged residues. This hydrophobic portion may traverse the cell membrane as two turns of α helix. A short, positively-charged section lies inside the cytoplasm.

Such interactions induce the production of the cytokines IL-1, IL-3, IL-6 and GM–CSF which are required for T-cell maturation in the thymus. Thymocytes also express receptors for IL-2 and this cytokine, together with other molecules, promotes cell proliferation which mainly occurs in the subcapsular and deep cortex.

Negative selection in the periphery (peripheral tolerance)

Not all self-reactive T cells are eliminated during intrathymic development, probably due to the inability of all self-antigens to transit through the thymic tissues. The thymic epithelial barrier may also limit access to some circulatory antigens. Given the survival of some self-reacting T cells, a separate mechanism is required to prevent their attacking the body. Recent experiments with transgenic mice have suggested that peripheral inactivation of self-reactive T cells can have two causes:

1. down-regulation of TCR and CD8 (in cytotoxic cells) so that the cells are unable to interact with target autoantigens
2. anergy, due to the lack of crucial secondary activation signals provided by the target cells.

Peripheral tolerance is discussed in more detail in Chapter 10.

Extrathymic T cell development

Although the vast majority of T cells require a functioning thymus for their differentiation, small numbers of cells (often oligoclonal in nature) carrying T cell markers have been found in athymic ('nude') mice. The possibility that there are thymic remnants cannot, however, be ruled out. There is some experimental evidence suggesting that bone marrow precursors can home to mucosal epithelia and then mature to functional TCR-1 T cells, without the need for a thymus.

B CELLS

In the chicken, primary B-cell lymphopoiesis occurs in a discrete lympho-epithelial organ, the bursa of Fabricius. The bursal rudiment develops as an outpushing of the hindgut endoderm and becomes seeded with blood-borne stem cells. Studies on chicken/quail chimeras have indicated that there is a window for the immigration of stem cells into the bursa between days 10 and 14 of embryonic life (see Chapter 14). Pyroninophylic cells – the putative stem cells – are seen in contact with epithelial cells. Bursal cell proliferation gives rise to the cortex and the medulla in each bursal follicle, which may be seeded by one or a few stem cells (Fig. 11.12).

Mammals do not have a specific discrete organ for B-cell lymphopoiesis. Instead, these cells develop directly from lymphoid stem cells in the haemopoietic tissue of the fetal liver (Fig. 11.13) from 8–9 weeks of gestation in humans, and by about 14 days in the mouse. Later the site of B-cell production moves from the liver to the bone marrow, where it is continued into adult life. This is also true of the other haemopoietic lineages, giving rise to erythrocytes, granulocytes, monocytes and platelets. Recent data have indicated that B-cell progenitors are also present in the omental tissue of murine and human fetuses. Whether or not these B-cell progenitors precede those in the fetal liver remains to be established.

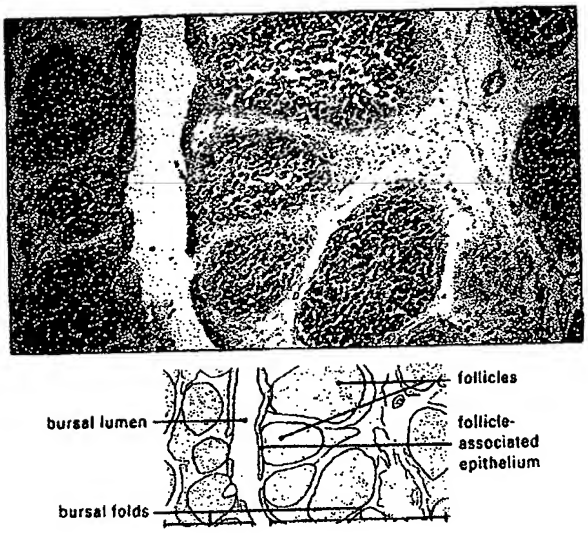


Fig. 11.12 Section of a bursa showing B cells developing in follicles. Like the fetal liver, the bursa is a site of some granulocytopoiesis, not just a site for lymphocyte development. H&E stain, ×50.

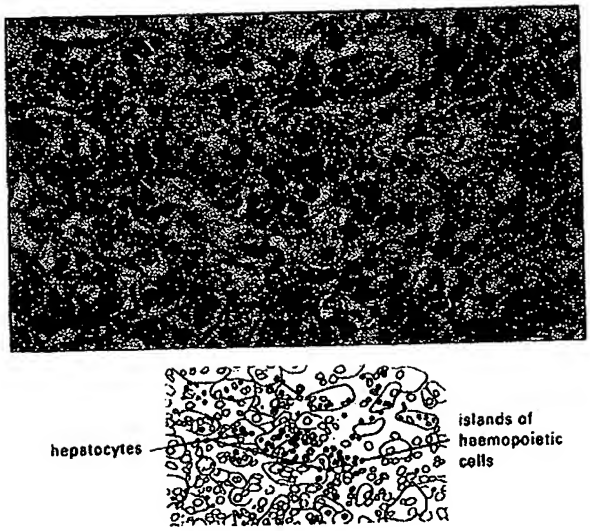


Fig. 11.13 Section of human fetal liver showing islands of haemopoiesis. Haemopoietic stem cells (HSCs) give rise to islands of differentiating lineage-specific cells, including B cells.

B-cell production in the bone marrow does not occur in distinct domains. However, it has been shown that B-cell progenitors are adjacent to the endosteum of the bone lamellae. Each B-cell progenitor, at the stage of immunoglobulin gene rearrangement, may produce as many as 64 progeny. These migrate towards the centre of each cavity of the spongy bone and reach the lumen of a venous sinusoid. In the bone marrow, B cells mature in close association with stromal reticular cells. The latter are found both adjacent to the endosteum and in close association with the central sinus, where they are termed adventitial reticular cells (Fig. 11.14). Reticular cells have mixed phenotypic features with some similarities to fibroblasts, endothelial cells and smooth muscle cells. They produce type IV collagen, laminin and the smooth-muscle form of actin. Experiments *in vitro* have shown that reticular cells sustain B-cell differentiation. Adventitial reticular cells may be important for the release of mature B cells into the central sinus. The majority of B cells (over 75%) maturing in the bone marrow do not reach the circulation but (as in the thymus) undergo a process of programmed cell death or apoptosis, and are phagocytosed by bone marrow macrophages. It has been suggested that B-cell–stromal interactions may mediate a form of positive selection that rescues a minority of B cells with productive rearrangement of their immunoglobulin genes from programmed cell death. Negative selection of autoreactive B cells may occur in the bone marrow or in the spleen, the site to which the majority of newly produced B cells are exported.

From kinetic data, it is estimated that about 5×10^7 murine B cells are produced per day. Since the mouse spleen contains approximately 7.5×10^7 B cells, a large proportion of B cells must die, probably at the pre-B-cell stage stage where they outnumber the B cells in the marrow by a factor of two.

The characteristic markers of the B-cell lineage are immunoglobulins, which act as cell-surface antigen receptors. Lymphoid stem cells (probably expressing terminal deoxynucleotidyl transferase, Tdt) proliferate and differentiate and then undergo immunoglobulin

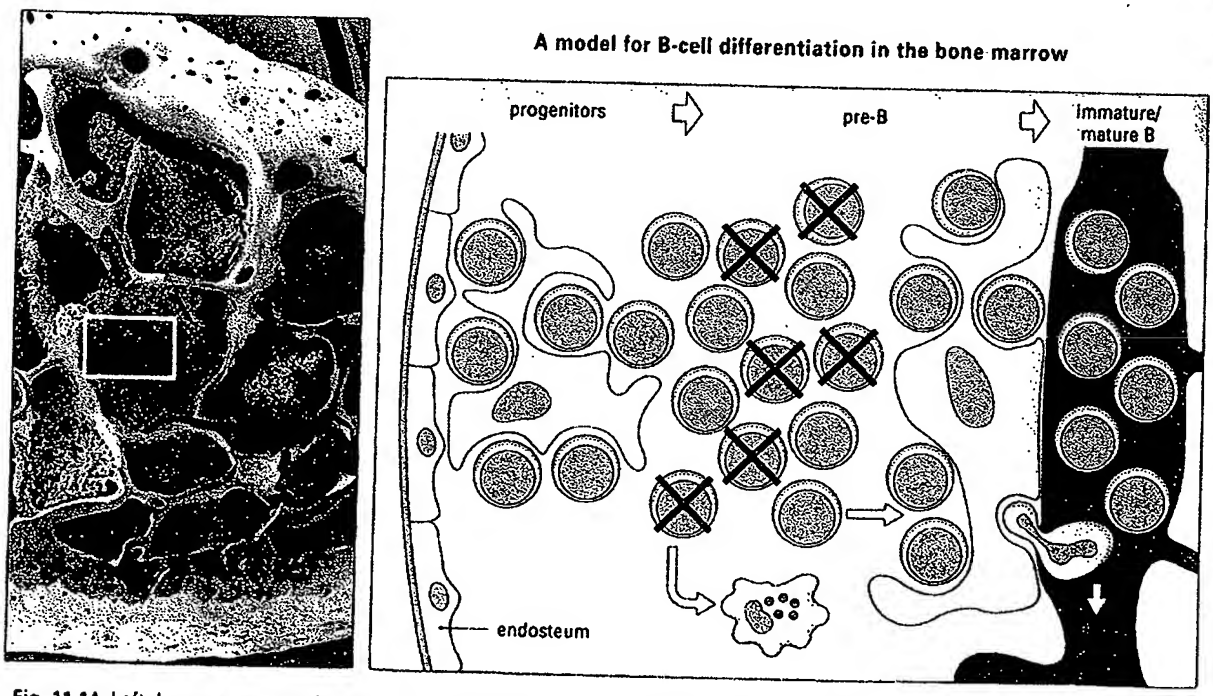


Fig. 11.14 Left: Low power scanning electron micrograph showing the architecture of bone and its relationship to bone marrow. A cavity has been picked out and is drawn schematically on the right. (Courtesy of Drs A. Stevens and J. Lowe). Right: Within the cavities of spongy bone, B cell lymphopoiesis takes place with maturation occurring in a radial direction towards the centre (from the endosteum to the central venous sinus). Immature progenitor cells adjacent to the endosteal cell layer mature into pre-B cells, most of which die and are phagocytosed by bone marrow macrophages containing tingible bodies (stained by haematoxylin). Cells which survive mature further and reach the central venous sinus. Association with reticular cells, and the presence of cytokines such as IL-7 is essential for all steps of B cell maturation. (Adapted from Osmond D, Gallagher R. *Immunol Today* 1991:12;1-3,).

gene rearrangements (see Chapter 4). Following this, they emerge as pre-B cells which express μ heavy chains in the cytoplasm. At least some of the pre-B cells express small amounts of surface μ chains associated with pseudo light chains VpreB and λ5. Allelic exclusion of either maternal or paternal immunoglobulin genes has already occurred by this time. The proliferating pre-B cells are thought to give rise to smaller pre-B cells. On synthesis of light chains, which may be either of κ or λ type, but not both, the B cell is then committed to the antigen- binding specificity of its sIgM antigen receptor. Thus one B cell can make only one specific antibody, a central tenet of the clonal selection theory for antibody production. A summary of B-cell differentiation, with expression of immunoglobulins and other relevant molecules, is shown in Fig. 11.15.

A sequence of immunoglobulin gene rearrangements and phenotypic changes takes place during B-cell ontogeny similar to that described for T cells (see above and Chapter 5). Heavy chain gene rearrangements occur in B-cell progenitors and represent the earliest indication of B-lineage commitment. This is followed by light chain gene rearrangements which occur at later pre-B-cell stages. Certain B-cell surface markers are expressed prior to immunoglobulin detection, namely class II MHC molecules, CD19, CD20, CD21 and the CD10 (CALLA) antigen. The latter marker is a highly conserved neutral endopeptidase which is transiently expressed on early B progenitors before the appearance of heavy μ chains in the cytoplasm. CALLA is re-expressed later in the B-cell life history, following activation by antigen (see Fig. 11.15). Other markers such as CD23 and CD25 (IL-2 receptor α) are mostly found on activated B cells.

A number of growth and differentiation factors are required to drive the B cells through early stages of development. Receptors for these factors are expressed at various stages of B cell differentiation. IL-7, IL-3 and low molecular weight B-cell growth factor (L–BCGF) are important in initiating the process of B-cell differentiation whereas other factors are active in the later stages (see Fig. 11.16).

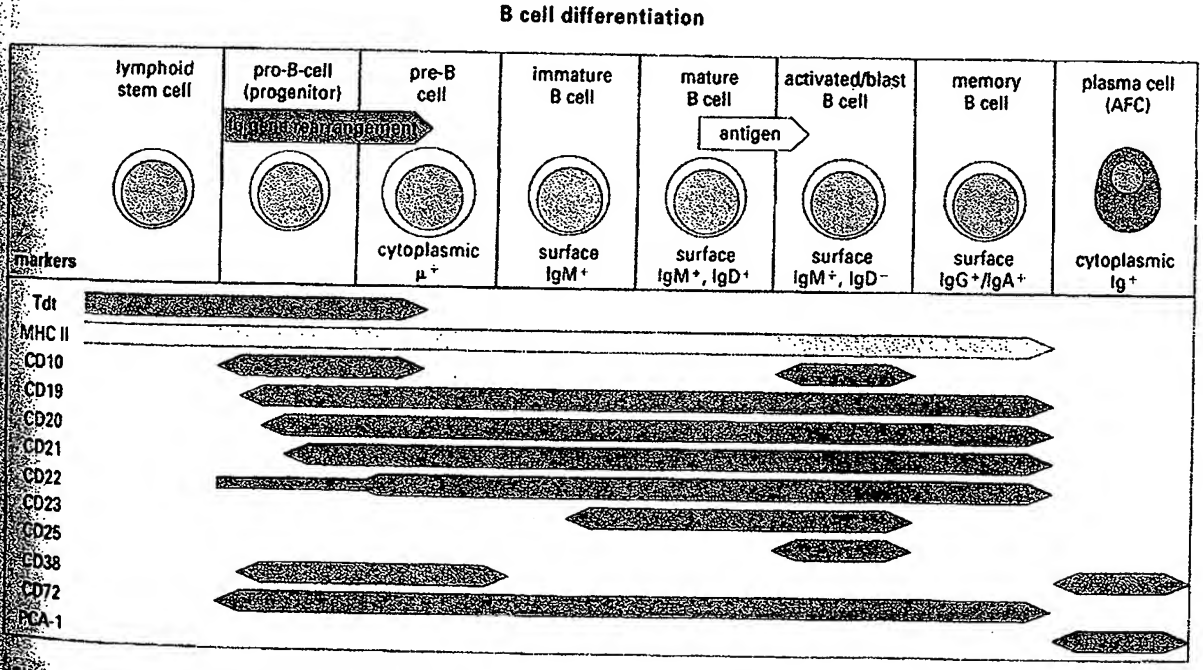


Fig. 11.15 B cells differentiate from lymphoid stem cells into virgin B cells and may then be driven by antigen to become memory cells or plasma cells. The cellular location of immunoglobulin is shown in yellow. The genes coding for antibody are rearranged in the course of progenitor cell development. Pre-B cells express cytoplasmic μ chains only. The immature B cell has surface IgM, and the mature B cell other immunoglobulin isotypes. On antigen stimulation the B cell proliferates and develops into a plasma cell or a memory cell following a phase of proliferation, activation and blast transformation. Memory cells and plasma cells are found at different sites in lymphoid tissue. Tdt is expressed very early in ontogeny. The diagram also shows the sequence of appearance of other important B cell surface markers. PCA-1 is found only on plasma cells. Note that CD38 is an example of a molecule found on early progenitors that is lost, only to reappear on the fully differentiated plasma cells.

Following their production in the fetal liver, B cells migrate and function in the secondary lymphoid tissue. Early immigrants into fetal lymph nodes (17 weeks in man) are $sIgM^+$ and carry a T-cell marker (CD5). $CD5^+$ B cell precursors are found in the fetal omentum. Small numbers of $CD5^+$ B cells are also found in the mantle zone of secondary follicles in adult lymph nodes.

Following antigenic stimulation, mature B cells can develop into memory cells or antibody-forming cells (AFCs). Surface immunoglobulin (sIg) is usually lost by the plasma cell (the terminally differentiated form of an AFC), since its function as a receptor is finished. Immature and mature B cells respond in different ways to antigens. Treatment with anti-IgM antibodies or antigen results in loss of sIgM by capping and endocytosis in both mature and immature B cells. However, only mature B cells resynthesize sIgM in culture (Fig. 11.17). Since immature B cells can be induced to lose their antigen receptor, this could be one mechanism by which self-reactive B cells are rendered tolerant during development.

DIVERSITY OF ANTIBODY SPECIFICITY

There are tens of thousands of natural antigenic shapes. Since one B cell can make only one antigen-specific antibody, many B cells with different specific antibody receptors have to be generated from the stem cells.

The genes encoding the variable regions of antibody molecules are described in detail in Chapter 5. The variable region genes comprising V, D and J segments are present in every somatic cell in a germ-line configuration. During early development intervening sequences between D and J are deleted, bringing these genes closer together. Further rearrangements of the V, D and J segments of the variable region heavy chain genes (V_H) occur during the progenitor stage of B-cell development (see Fig. 11.15). The various combinations of these genes (one productive rearrangement per cell) are expressed with μ heavy chain genes in the cytoplasm of the large pre-B cell. These actively proliferating pre-B cells then rearrange their $V\kappa$ genes, and later their $V\lambda$ genes, if the κ rearrangement has not been successful. When a light chain gene is productively rearranged the immature B cell expresses surface μ chains with either κ or λ light chains. Those cells not making productive rearrangements probably die (perhaps by apoptosis). This is one explanation as to why so many pre-B cells die during development (see above). There is some evidence that pseudo-light-chain genes are expressed prior to κ and λ light chains, and that these may assemble small amounts of surface IgM on pre-B cells. This might be important in selection of early pre-B cells.

Once the κ or λ light chains are being produced, the surface IgM on the immature B cell can act as a functional antigen receptor. The rearrangements of the V, D and J segments (heavy chains) and V and J segments (light chains) are thought to be randomly generated within the B cells. However, there is evidence in mice, rats and chickens for a programmed sequence of development of specific antibody specificities (Fig. 11.18). Antibody production, as distinct from antigen recognition by B cells is, however, dependent on both T cells and APCs.

The reason for this programmed development of specificities at the molecular level within B cells is unclear, but it may reflect the biased utilization of V gene segments nearest to the D or J segments, with a 3' directional movement of the relevant recombinases and/or some negative selection of particular clones (possibly for self-reactivities). Many of the first B cells to appear in ontogeny express a predominantly T cell-

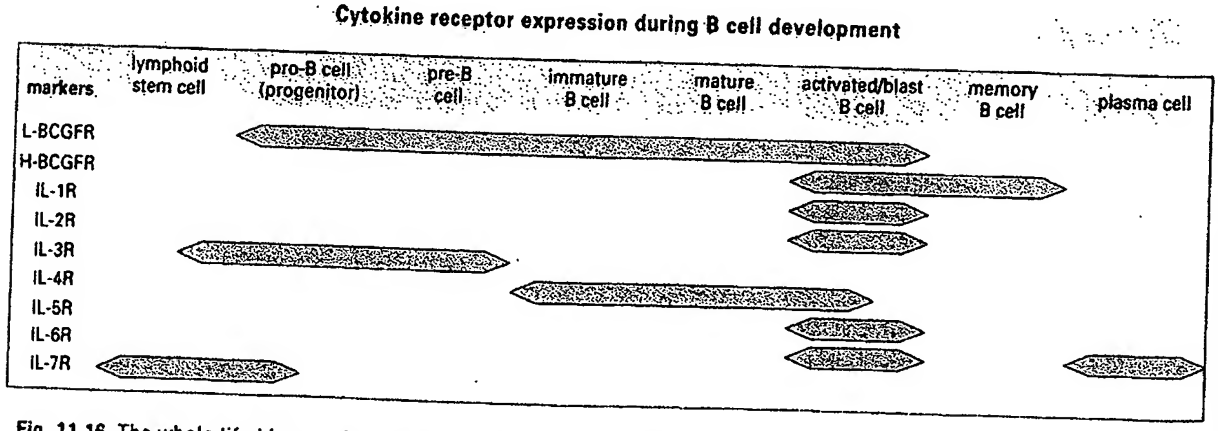


Fig. 11.16 The whole life history of B cells from stem cell to mature plasma cell is regulated by cytokines present in their environment. Receptors for these cytokines are selectively expressed by B cells at different stages of development. IL-7 plays an important role in initiating events in B cell differentiation.

Immunogenetics (1997) 45: 386–393

ORIGINAL PAPER

Annex II

Mark D. Schrenzel · Donald P. King · Mark L. McKnight
David A. Ferrick

Characterization of horse (*Equus caballus*) immunoglobulin mu chain-encoding genes

Received: 14 October 1996 / Revised: 10 December 1996

Abstract Horse (*Equus caballus*) immunoglobulin mu chain-encoding (*IgM*) variable, joining, and constant gene segments were cloned and characterized. Nucleotide sequence analyses of 15 cDNA clones from a mesenteric lymph node library identified 7 unique variable gene segments, 5 separate joining segments, and a single constant region. Based on comparison with human sequences, horse variable segments could be grouped into either family 1 of immunoglobulin (*Ig*) clan I or family 4 of *Ig* clan II subclan IV. All horse sequences had a relatively conserved 16 base pair (bp) segment in framework 3 which was recognized with high specificity in polymerase chain reaction by a degenerate oligonucleotide primer. Horse complementarity determining regions (CDR) had considerable variability in predicted amino acid content and length but also included the presence of relatively conserved residues and several canonical sequences that may be necessary in formation of the β chain main structure and conformation of antigen-binding sites through interaction with light chain CDR. Sequence analysis of joining regions revealed the presence of nearly invariant 3′ regions similar to those found in human and mouse genes. A single horse *IgM* constant region comprising 1472 bp and encoding 451 residues was also identified. Direct comparison of the horse constant region predicted amino acid sequence with those from eleven other species revealed the presence of 53 invariant residues with particularly conserved sequences within the third and fourth exons. Phylogenetic analysis using a neighbor-joining algorithm showed closest similarity of the horse mu chain-encoding constant region gene to human and dog sequences. Together, these findings provide insights into the comparative biology of *IgM* as well as data for additional detailed studies of the horse immune system and investigation of immune-related diseases.

The nucleotide sequence data reported in this paper have been submitted to GSDB and have been assigned the accession numbers L81155 (clone Ig1); L81156 (clone Ig51); L81157 (clone I14); L81158 (clone I16); L81159 (clone I18); L81160 (clone Ig19); L81161 (clone Ig20), and L49414 (mu chain-encoding constant region)

M. D. Schrenzel (✉) · D. P. King · D. A. Ferrick
Department of Pathology, Microbiology, Immunology, School of Veterinary Medicine, University of California, Davis, CA 95616, USA

M. L. McKnight
Section of Evolution and Ecology, Center for Population Biology, University of California, Davis, CA 95616, USA

Introduction

Immunoglobulins (*Ig*) constitute the major functional protein of humoral immunity and are comprised of heavy and light chains, each containing variable and constant genes (Honjo 1983; Chothia and Lesk 1987; Tutter and Riblet 1989). *Ig* heavy chain variable domains are generated by the recombination of discontinuous germline variable (*Igh-V*), diversity (*Igh-D*), and joining (*Igh-J*) segments. Within the *Igh-V* molecule, differences in peptide sequences are clustered in three intervals, termed hypervariable or complementarity determining regions (CDR), which are separated by conserved framework regions (FW). The CDR comprise the antigen-binding portions of immunoglobulins, while the FW provide structural stability to allow proper association of light and heavy chain heterodimers (Kirkham et al. 1992). The heavy chain constant genes define the different *Ig* subclasses as well as aspects of their immunologic functions and have been named *IgM*, *IgD*, *IgG*, *IgA*, and *IgE*. Among these, the IgM molecule appears to be the most primitive, since it was the first to appear in vertebrate evolution (phylogeny) and is the first Ig expressed during ontogeny of the immune system (Bosch et al. 1992).

In humans and mice, *Igh-V*, *Igh-D*, *Igh-J* and *Igh-M* constant genes have been extensively characterized with similar but less detailed work reported for rats, sheep, cattle, dogs, cats, rabbits, hamsters, shrews, pigs, ducks, and several fish (Bosch et al. 1992; Meek et al. 1990; Parker et al. 1994; Vasquez et al. 1992; Hordvik et al. 1992; Ishiguro et al. 1989; McCumber and Capra 1978; Bernstein et al. 1982; Patri and Nau 1992; McGuire et al. 1985). Based on comparative analyses from these studies, mammalian variable gene segments appear to have been derived

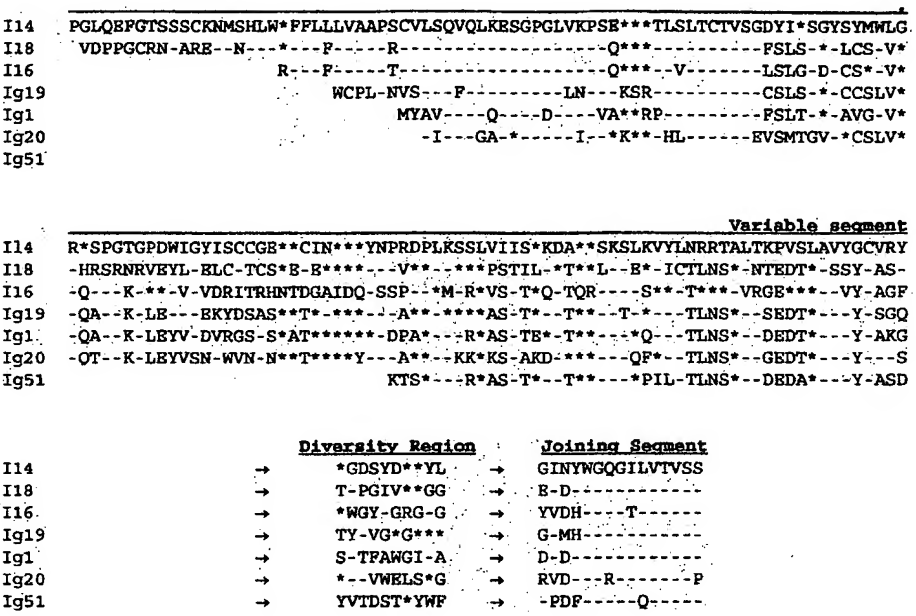


Fig. 1 Alignment of predicted amino acid sequence for seven cDNA clones encoding horse *Igh-V, Igh-D* and *Igh-J* genes. The ends of the variable gene segment and the diversity segment are estimates based on lineups with human and mouse germline *Ig* sequences. Sequence identity with clone I14 are shown as *dashes* (–). Gaps (*) were inserted to maximize alignment. Nucleotide sequences of cDNA clones have been submitted to GSDB (accession numbers L81155-L81161)

from three distinct progenitor genes (Tutter and Riblet 1989). Descendants from these genes are categorized into three clans based on nucleotide sequence similarity in FW1 and FW3 regions (Kirkham et al. 1992). Each clan contains one or several families (>80% nucleic acid sequence identity places two genes within the same *VH* family (Cook and Tomlinson 1995). In humans, at least 95 *Igh-V* segments comprising 7 families and 6 *Igh-J* gene segments have been identified (Chothia and Lesk 1987; Cook and Tomlinson 1995). In mice, 13 *Igh-V* families and 4 *Igh-J* segments are known (Meek et al. 1990). In horses, only the IgE heavy chain, including four *Igh-V* sequences, have been characterized (Navarro et al. 1995, and unpublished data GSDB accession number: U17041). In this study, we characterize the nucleotide sequences of 7 unique horse *Igh-V* genes, 5 *Igh-J* segments, and the *IgM* constant region, including the secretory peptide, and compare our findings with those from other species.

Materials and methods

Animals

Samples for RNA and DNA extractions were provided by the Equine Research Laboratory and Veterinary Medical Teaching Hospital at the University of California, Davis. RNA for PCR detection of heavy chain rearrangement and clonality expression was obtained from normal mesenteric lymph node from a healthy adult Thoroughbred horse and an adult Quarterhorse with lymphosarcoma. DNA for Southern blot hybridizations was obtained from kidney.

Polymerase chain reaction (PCR)

RNA was isolated from mesenteric lymph node from a healthy two-year-old Thoroughbred horse and from a lymphoid tumor from a six-year-old Quarterhorse using the RNAzol B method (TM CinnaScientific, Friendswood, TX) according to the manufacturer's instructions. Seven µg of total RNA was used for cDNA synthesis with MMLV reverse transcriptase (Gibco BRL, Eggenstein, Germany) and random hexanucleotides as described (Schrenzel et al. 1994). PCR reactions contained 5 mM $MgCl_2$, 50 mM KCl, 10 mM Tris-HCl pH 8.0, 0.2 mM of each dNTP, 50 pmol of each primer, and 0.5 units *Taq* DNA polymerase (Promega, Madison, WI) in a final volume of 50 µl. Primers used for detection of horse *IgM* gene expression were degenerate and based on alignments of mu chain-encoding constant regions of other species. The sense primer was AAG CC(AC) G(AT)G GCC TGG (CT)TG TC and the anti-sense primer CTA G(AG)C TGC TGT GCT TCA CCC designed to generate a 290 bp PCR product. PCR conditions were 95 °C for 45 s, 54 °C for 1 min, and 72 °C for 1 min for 30 cycles. This PCR fragment was cloned into pGEM-T using the TA Cloning Kit from Promega (Madison, WI) according to the manufacturer's instructions. Primers used to amplify a broad spectrum of horse *Ig* genes were a degenerate oligonucleotide primer GGCC(GAC)(CT)(GAC)TATT(AT)CTGT designed to bind FW3 based on alignments of horse *Igh-V* genes and an anti-sense constant region primer GTGTTGTTCTGGTAGTTCCAG. PCR conditions were 95 °C for 45 s, 54 °C for 1 min, 72 °C for 1 min for 26 cycles. cDNA from normal horse lymph node and from a horse B-cell lymphosarcoma along with plasmid-cloned horse *Ig* genes were used in the PCR.

cDNA library construction and screening

Total RNA from mesenteric lymph nodes of five healthy unrelated horses (three Thoroughbreds, one Standardbred, one Quarterhorse) was isolated over cesium chloride gradient and used by Stratagene (La Jolla, CA) to construct a size-fractionated lambdaZapII cDNA library with oligo-dT primers. The cDNA library was screened using a horse-specific 290 bp *IgM* constant region probe (described above) labeled with [alpha-P32]dCTP using a hexanucleotide random priming system (Boehringer Mannheim, Indianapolis, IN). Plaque hybridization was carried out for 18 h at 60 °C in 5× standard sodium citrate (SSC), 1× Denhardt's solution, 0.1% sodium dodecyl sulfate (SDS), and 100 µg/ml denatured herring sperm DNA. Filters were washed twice with 2× SSC and once with 1× SSC at 60 °C. Single clones were obtained by secondary and tertiary screenings under the same conditions. Inserts were subcloned into the Bluescript phagemid vector by in vivo excision with Ex Assist helper phage (Stratagene).

Gene manipulation and DNA sequencing

Cloned PCR products and isolated cDNA library clones were sequenced by the chain termination method (Sanger et al. 1977) with Sequenase 2.0 (US Biochemicals, Cleveland, OH) and with a Prism automated sequencing machine (Perkin Elmer, Norwalk, CT) with the M40 forward primer, M13 reverse primer, and primers designed for regions of the horse *IgM* constant region gene.

Sequence analyses

Sequences were assembled and analyzed with the University of Wisconsin Genetics Group (GCG) computer program. A neighbor-joining tree showing a hypothesis of the evolution of the *IgM* constant region gene was produced by the Phylogenetic Analysis Using Parsimony (PAUP) program and the data were bootstrapped to determine levels of support (Saitou and Nei 1987; Swofford 1993).

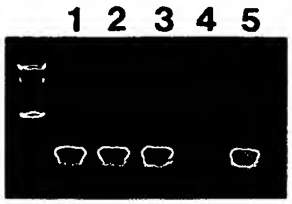


Fig. 2 PCR products obtained from amplification with sense degenerate oligonucleotide primer for conserved *Ig* FW3 region and antisense primer to *IgM* constant region, stained with ethidium bromide, and run on 2% agarose gel; 100 bp ladder with double bright band representing 600 bp on *left*; Lanes 1–3 selected plasmids containing horse *Ig* genes; Lane 4 cDNA from normal horse lymph node; Lane 5 horse B-cell lymphosarcoma. Note discrete bands for plasmid clones and monoclonal B-cell lymphosarcoma but different size PCR products from lymph node resulting in broad smear

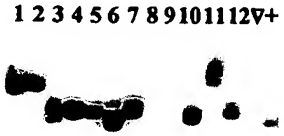


Fig. 3 PCR products obtained from amplification with sense degenerate oligonucleotide primer for conserved *Ig* FW3 region and antisense primer to *IgM* constant region, labeled with P32-dCTP, and run on 6% polyacrylamide gel; Lanes 1–7 and 10–12 horse *IgM* samples cloned into Bluescript plasmid; Lanes 8 and 9 horse *Tcr* beta chain-encoding and *Tcr* delta chain-encoding genes cloned into Bluescript plasmid; Lane inverted triangle (∂) cDNA from normal horse lymph node; Lane plus (+) cDNA from horse B-cell lymphosarcoma. Note lack of band formation for *Tcr* genes and polyclonal lymph node cDNA but specific recognition of all plasmids containing horse *Ig* genes and monoclonal B-cell lymphosarcoma

Results

Fifteen clones, including seven unique *Igh-V* segments, were isolated from a horse mesenteric lymph node cDNA library using a 290 bp PCR fragment probe to the *IgM* constant region. Five of the seven variable segments were full-length, and all clones contained FW3 (Fig. 1). Nucleotide sequence analyses for horse *Igh-V* regions showed closest sequence identity with other published horse *Igh-V* segments followed by sheep, cattle, and human genes (data not shown). Comparison of the horse *Igh-V* segments with human sequences revealed closest overall nucleotide similarity with either family 1 of Clan I or family 4 of Clan II when using the GCG gap comparison (data not shown). Comparison of the seven horse *Igh-V* genes with each other similarly demonstrated the presence of two families. However, all horse *Igh-V* genes had relatively conserved FW3 regions. The horse FW3 regions were similar in size to those found in other species and allowed design of a single degenerate primer for recognition of all seven *Igh-V* genes. This primer was used in the polymerase chain reaction (PCR) with an anti-sense oligonucleotide primer to the constant region and amplified the seven cloned horse *Igh-V* genes as well as cDNA from a normal horse lymph node and a B-cell lymphosarcoma (Figs. 2, 3). FW1 regions from the five full-length horse *Igh-V* genes were also highly conserved.

Variability analysis of horse *Igh-V* genes demonstrated the presence of three CDR (Figs. 1, 4). CDR1 (residues 26–35) had a constant length of 10 amino acids except for two sequences (I14 and I16) which had additional insertions of two residues. Among the horse *Igh-V* genes, 4 of 6 had unique CDR1 sequences when compared with data compiled from 77 human *IGH-V* gene sequences, but all had an invariant valine residue at the adjoining residue 24 position and 5 of 6 had a glycine at position 26. Horse CDR2 (residues 50–65) had considerable variability but did contain a relatively conserved asparagine residue at position 61. Horse CDR3 (residues 96–101) varied significantly in residue content and length with sequences containing from 6–10 CDR3 amino acids. However, 5 of 6 sequences contained a leucine or phenylalanine near residue 100.

Horse *Igh-D* regions were estimated based on comparisons with germline *Igh-V* and *Igh-J* sequences from other species and contained from 6–10 amino acids with an average of 8.1. Among the CDR, the regions formed within the *Igh-V*, *Igh-D*, and *Igh-J* junctional region had the highest level of variability. However, 6 of 7 horse *Igh-D* regions contained at least one glycine residue within the *Igh-D* junctional region and 5 of 7 contained a tyrosine residue (Fig. 1).

Seven different horse *Igh-J* genes were identified. However, three of these genes differed by only 1 or 2 bp and, thus, may represent alleles of the same gene or reflect differences in N region modification at the 5' terminus of the same gene segment. Thus, our analyses appear to have identified five unique *Igh-J* segments (Fig. 1). The five separate horse *Igh-J* genes had from 82.5% to 94.0% nucleotide identity with each other and an average of 90.3% identity. Predicted amino acid sequences for the 7 horse *Igh-J* sequences demonstrated highly conserved areas, particularly at the 3' terminus, although the 5' terminus in five of seven horse *Igh-J* genes contained a tyrosine residue.

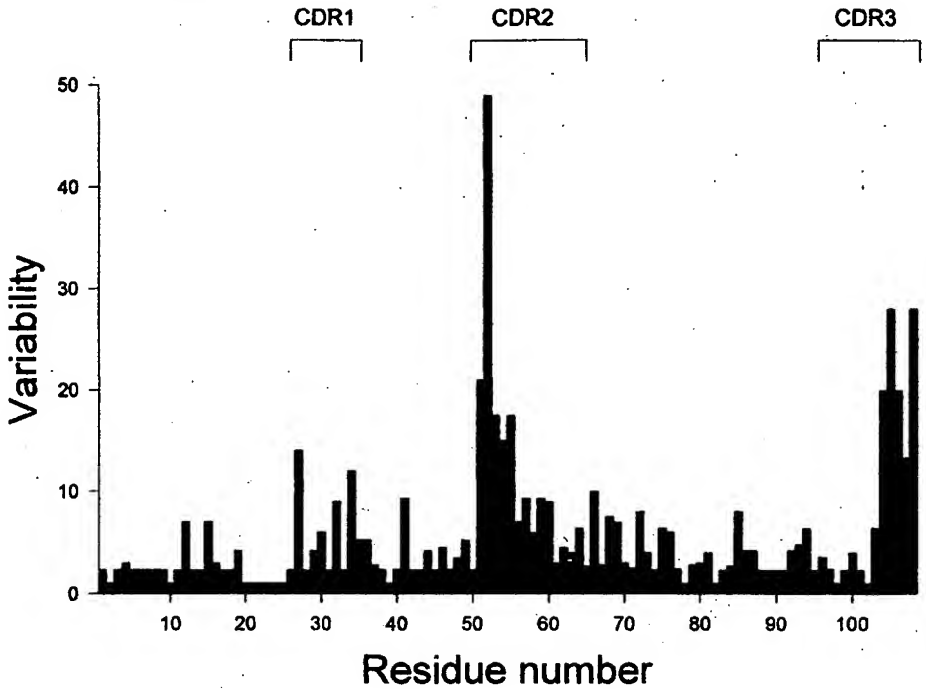


Fig. 4 Variability analysis of seven translated *Igh-V* gene segments. Variability was calculated as the number of different amino acids occurring at a given position divided by the frequency of the most common amino acid at that position (Kabat 1987). Approximations of the complementarity determining regions (CDR) are shown above and were made based on comparison with human sequences (Chothia and Lesk 1987; Chothia et al. 1992)

Among the 15 horse genes sequenced, all had the same *IgM* constant region consisting of a 1472 bp segment encoding for 451 amino acids with six potential *N*-glycosylation sites (Fig. 5). Also present was a 58 bp segment coding for 19 amino acids that represented the IgM secretory peptide. Southern blot hybridization of the horse *IgM* constant gene to genomic DNA revealed the presence of a single band for *Bam* HI and *Eco* RI restriction enzyme digests (data not shown). A lineup comparison with ten mammalian species, including horse, and two non-mammalian species (duck and salmon) demonstrated the presence of 53 invariant residues and 159 highly conserved (present in at least 9 of 12 species) amino acids throughout the length of the *IgM* constant region (Fig. 6). Nucleotide and amino acid sequence comparisons of the horse *IgM* constant region with other species showed closest similarity to human and least similarity to duck and salmon. Among the four mu chain-encoding exons, CH3 and CH4 were the most highly conserved between species. A phylogenic tree calculated using neighbor-joining analysis is shown in Figure 6. In addition, parsimony analysis revealed similar phylogenetic groupings represented by high bootstrap values in the neighbor-joining tree and lack of clear resolution between other groups (data not shown).

Discussion

Ig architecture accommodates an abundance of structural diversity allowing binding of numerous antigens within the confines of a conserved framework. Conservatism among Ig molecules with different antigen affinities was explained by discoveries of somatic recombination and junctional diversification for most mammals and intergenic conversion in birds and rabbits. These findings were based on landmark analyses of *Ig* nucleotide sequences in mice, humans, chickens, and rabbits (Honjo 1983; Tonegawa 1983; Thompson 1992). Subsequently, *Ig* gene sequences have been reported for at least 18 species, including fish and amphibians. In this report, we characterize seven horse *IgM* cDNA clones and compare the predicted amino acid structures of variable, diversity, joining, and constant region segments with those from other species to better understand the comparative biology of *Ig*.

All seven of our horse *Igh-V* sequences had highest nucleotide identity with one of three previously characterized horse *Igh-V* sequences, and four had greater than 80% identity. These findings indicate that certain *Igh-V* families may be highly represented in horses. In humans, VH3 *Ig* gene segments from Clan III are most highly represented in both the germline and peripheral repertoire and are considered to be essential and conserved components of the *Ig* loci in a variety of mammalian and non-mammalian species (Cook and Tomlinson 1995; Tutter and Riblet 1989). The Clan III *Igh-V* genes are also preferentially expressed during fetal life and may function in recognition of primitive, conserved antigens (Kirkham et al. 1992). When compared with other species, horse *Igh-V* genes had highest nucleotide similarity to sheep, cattle, and human genes and least similarity to mouse and rat sequences. These findings were not unexpected, since horses have shared habitats with other herbivores, and thus have been exposed to similar selection pressures that shape development of the peripheral immunoglobulin repertoire. Additionally, artiodactyls

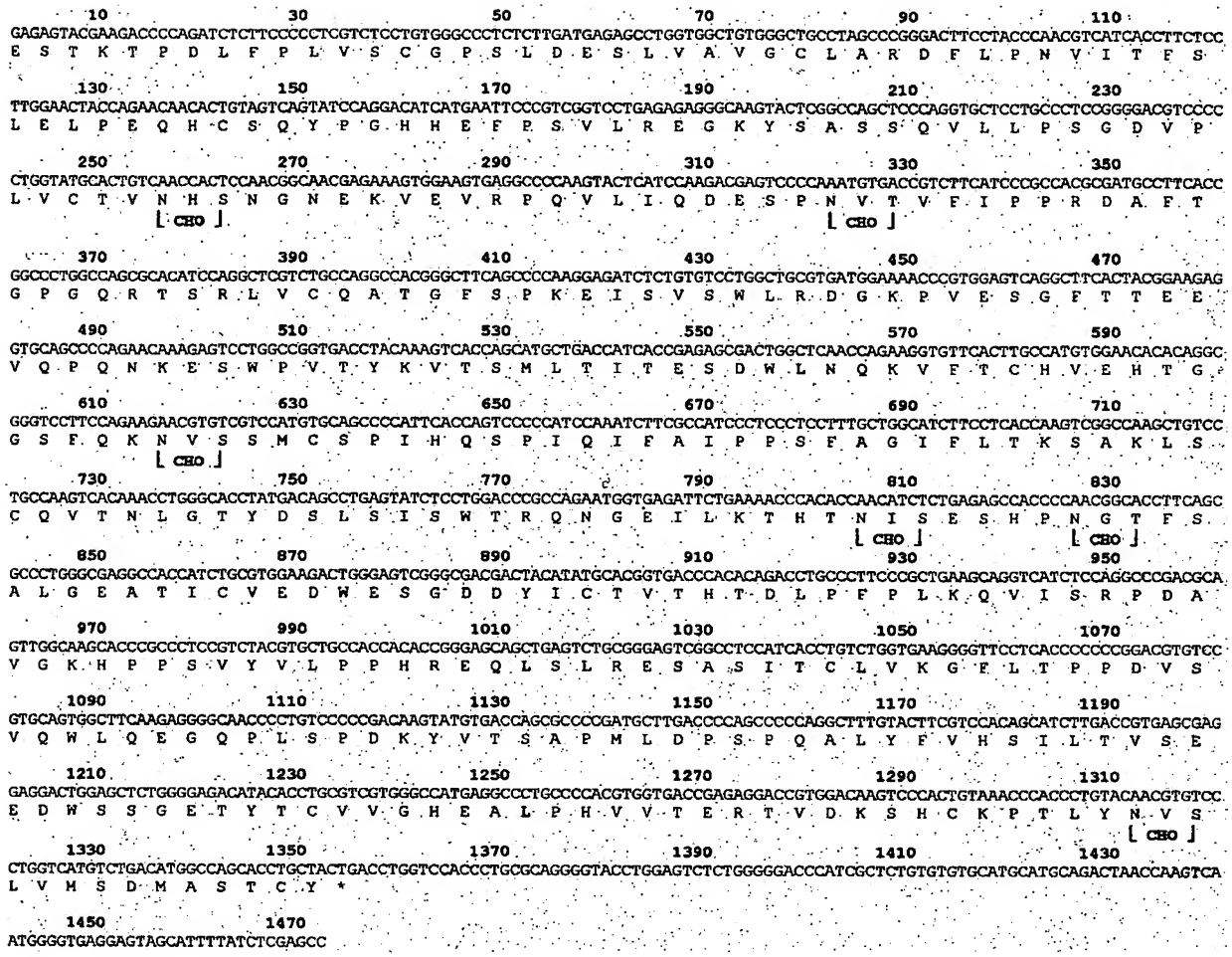


Fig. 5 Complete nucleotide sequence horse *IgM* constant region gene. Predicted amino acid sequence for the translated protein and potential *N*-glycosylation sites are shown

and persidactyls probably shared a more recent common ancestor than carnivores, rodents, and primates (Novacek 1992). The high similarity of horse *Igh-V* genes to sheep and cattle genes is similar to findings for horse *Tcrd-V* which were most closely related to sheep sequences (Schrenzel and Ferrick 1995). The similarity of some horse *Igh-V* with human genes may have resulted from common exposure to certain pathogens or other exogenous antigens but more likely represents an artifact stemming from the more numerous human *Igh-V* sequences in the genebank.

Variability plot analyses of horse *Igh-V* predicted amino acid sequences demonstrated the presence of CDR and FW similar to those of other species (Cook and Tomlinson 1995; Chothia et al. 1992; Chothia and Lesk 1987). A significant finding was the presence of an 18 bp conserved region within FW3 recognized with high specificity by a single degenerate oligonucleotide primer in all horse *Igh-V* clones, a normal lymph node, and a B-cell lymphosarcoma. These results are similar to those for human *IGH-V* for which a single universal variable gene segment primer has been identified. This primer is believed to recognize approximately 80% of all human *IGH-V* genes, including those in different Clans, and has proven to be a valuable diagnostic reagent for assessing B-cell clonality in lymphoid lesions (Ben-Ezra 1992; Wan et al. 1990). Like the human and mouse, horse *Igh-V* sequences also had highly conserved residues within CDR1 and CDR2 (Chothia et al. 1992; Chothia and Lesk 1987; Cook and Tomlinson 1995). These conserved residues were a valine at position 24 and a glycine at position 26 for CDR1 and an asparagine residue at position 61 for CDR2. These amino acids are believed to be important in maintaining the three-dimensional packing structure associated with formation of antigen-binding sites within the Ig molecules (Chothia and Lesk 1987). However, when horse *Igh-V* CDR1 sequences are compared with the human CDR1 canonical sequences following positions 24 and 26, they differ significantly. The human CDR1 canonical sequences are believed to be primarily responsible for the reliable duplication of main-chain structures present in antigen-binding sites by influencing packing, hydrogen bonding, or the ability to assume unusual φ, ψ, or ω conformations (Chothia and Lesk 1987). The finding of unique CDR1 sequences in horse *Igh-V* suggest that a

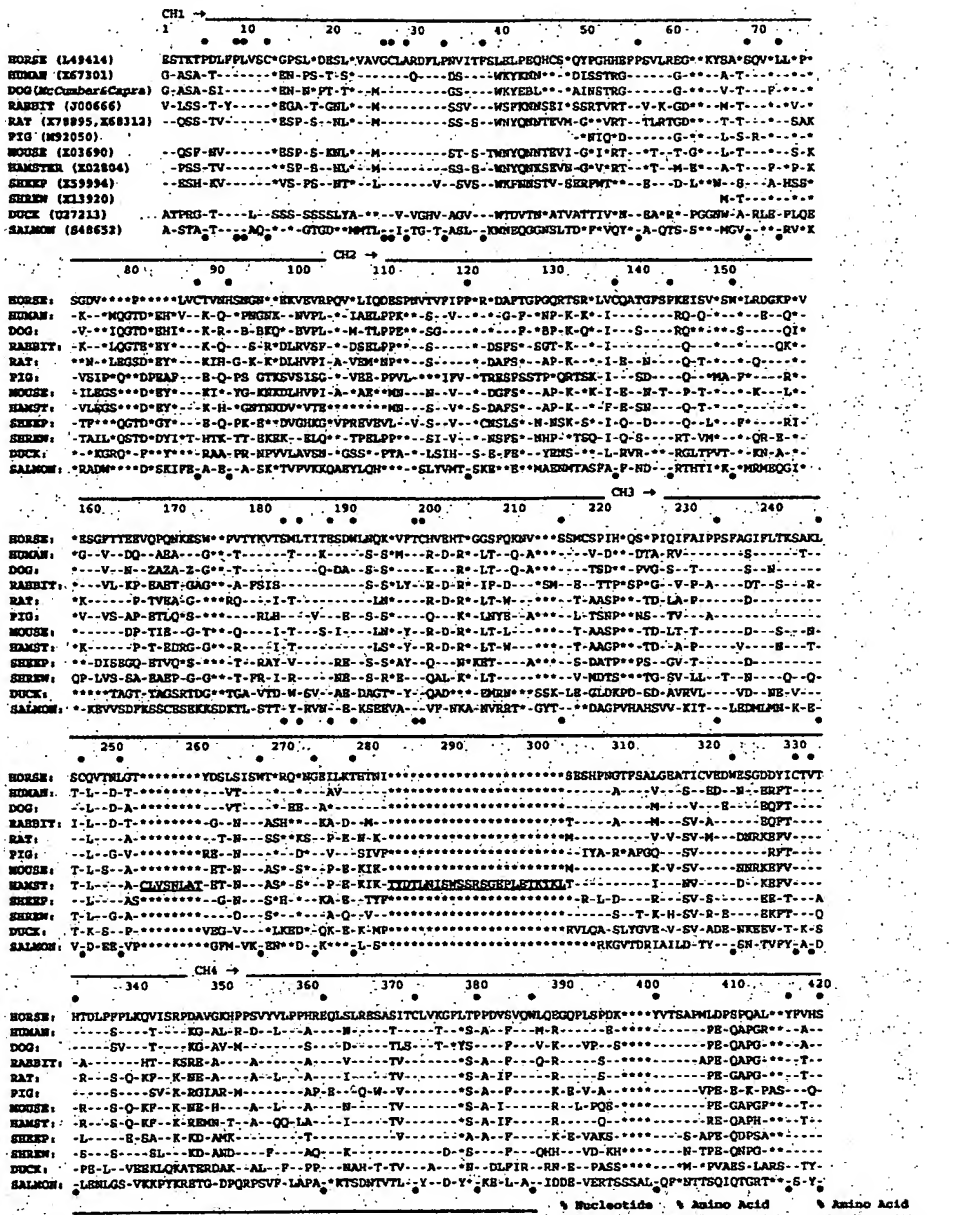


Fig. 6 Alignment of horse *IgM* constant region predicted protein sequence with other species. *Dashes* (–) represent amino acid residues conserved with the horse sequence. *Black circles* indicate invariant residues across all species shown. Gaps (*) were inserted to maximize alignment. A repetitive peptide motif identified in the hamster gene is *underlined*. Percentage identities and similarities of horse *IgM* constant region with other species is shown at the *bottom right*. References with accession numbers are: human (X67301, Harindranath et al. 1993); dog (McCumber and Capra 1978); rabbit (J00666, Bernstein et al. 1982); rat (X78895, Parker et al. 1994); pig (M92050, Bosch et al. 1992); hamster and mouse (X02804 and X03690, McGuire et al. 1985); sheep (X59994, Patri and Nau 1992); shrew (X13920, Ishiguro et al. 1989); duck (U27213, unpublished), and salmon (S48652, Horkvik et al. 1992)

different structural motif may contribute to formation of the antigen-binding area. Like the human and mouse CDR3, the horse CDR3 contained a phenylalanine or leucine in most sequences. Additionally, a glycine was present in six of seven horse *Igh-D* regions, indicating the importance of this residue in the structural integrity of *Ig* across species.

Based on sequence comparisons, at least five distinct horse *Igh-J* gene segments appear to exist. This is not unexpected, since humans have six joining gene segments and mice contain four. The presence of additional horse *Igh-J* segments, however, cannot be excluded without more extensive gene analyses. The presence of significant vari-

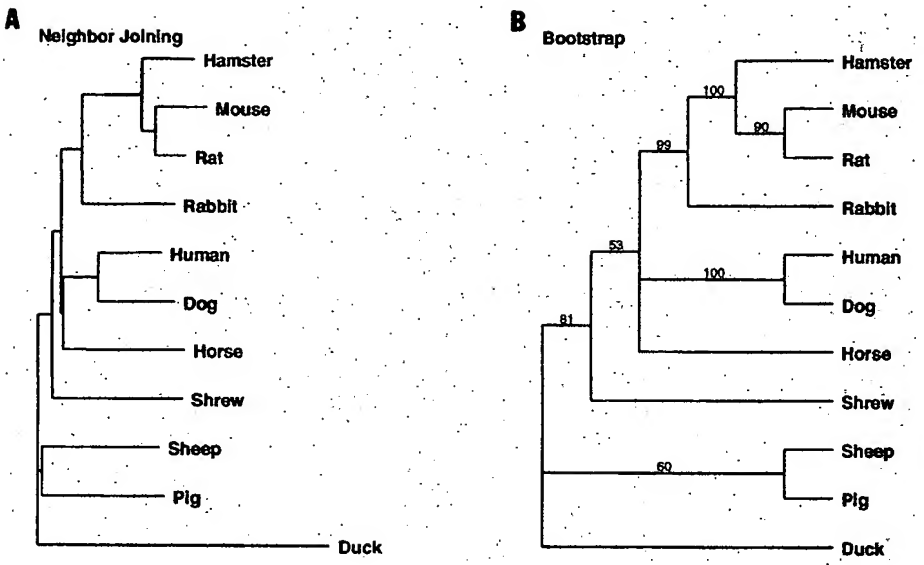


Fig. 7 A Neighbor-joining phylogram representing genetic distance of published *IgM* constant region genes. The tree was drawn using sequenced or predicted amino acid residues such that lengths of the horizontal branches are proportional to the number of amino acid substitutions that occurred on those branches. For references see Figure 6. **B** The bootstrap tree represents 1000 cycles of random character sampling with replacement (Felsenstein 1985). The bootstrap values indicate the level of support for the numbered branches

ability in the 5' region of the horse *Igh-J* genes indicated that, like mice and humans, junctional diversification occurs through modification of the joining segment as well as the diversity and variable genes. It is also likely that further modification the 5' horse *Igh-J* genes occurs due to somatic hypermutation as in other species (Chothia and Lesk 1987). Despite diversification of the 5' joining segment and its functional role as a constituent of the hypervariable region, 5/7 horse *Igh-J* genes sequenced contained a tyrosine residue in this region. This is similar to findings for humans and mice which have a 5' tyrosine in 4/6 and 4/4 joining segments, respectively (Chothia and Lesk 1987). The significance of these observations is not certain, but it has been shown that tyrosine residues have a high normalized frequency of occurrence in β-sheets (Creighton 1984). Phenylalanine, another aromatic amino acid, likewise occurs frequently in β-sheets and is present in 5/6 human *IGH-J* and 3/4 mouse *Igh-J* in the 5' region (Creighton 1984; Chothia and Lesk 1987). Thus, it appears that the presence of an aromatic amino acid in the 5' joining segment is important to the integrity of the *Ig* structural domain in this area, perhaps by contributing to β-sheet formation.

A single horse *IgM* constant region was present and contained highly conserved cysteine and tryptophan residues as seen in other species. Additionally, comparison of horse *IgM* sequence with nine other mammals and two non-mammalian species revealed the presence of 53 invariant residues and 159 highly conserved amino acids. The high level of structural similarity within the constant region may be related to the constraints placed on the constant region for assembly of light chains and a duplicate heavy chain to form a functional protein. Conversely, the conservation of amino acids, especially within exon 4, may be a reflection of the primary role of IgM in activation of the complement system which is highly conserved even between divergent species, the role of IgM as an Fc-bound surface receptor, or the requirement for polymerization of IgM molecules (McCumber and Capra 1978). Among the four exons comprising the constant region, exon 4 by far contains the most invariant and conserved residues. All twelve species analyzed also contained a potential *N*-glycosylation site and a cysteine residue within the secretory peptide. These regions have been suggested to be critical to the formation of IgM multimers and, in accordance with previous reports, indicate that multimers of secreted IgM likely occur in birds and fish as in mammals. The horse, like other mammals, lacks other potential *N*-glycosylation sites within exon 4. In contrast, many lower vertebrates contain one or several *N*-glycosylation sites in the last coding exon. The significance of this potential for structural diversity is unclear but may be related to the role of IgM in complement activation or possibly multimer formation. An additional feature that was revealed after alignment of all species was the presence of a repetitive 15 amino acid motif within exon 3 in the hamster gene. The presence of this repetitive sequence is similar to findings for the sheep and cattle *Tcrg-C* genes which contain duplicated peptide motifs within the connecting peptide (hinge region; Hein et al. 1990; Takeuchi et al. 1992). The functional significance of these repetitive sequences has never been determined. Unlike most other Ig isotypes and T-cell receptor proteins, IgM does not contain a hinge region (Kuby 1994).

Comparison of the horse constant region amino acid sequence with other species using a phylogenetic analysis indicated that the horse *IgM* chain has closest evolutionary distance to humans and dogs. This result was somewhat unexpected, particularly since the horse *Igh-V* genes were most closely related to the artiodactyls (sheep and cattle). Interestingly, the horse *Tcrb-C* and *Tcrd-C* also have highest similarity to human and dog sequences (Schrenzel et al. 1994; Schrenzel and Ferrick 1995). The significance of these findings is unclear. Perhaps, structural conservation of the T-cell receptor and *Ig* constant regions among horse, human, and dog indicates a relatively close evolutionary relationship between the immune systems of these species

on a molecular level. Analysis of other immune cell proteins, such as Thy-1 production by T-cells, also demonstrates striking similarities between horses and humans (M.D. Schrenzel, unpublished observation). Furthermore, a recent report comparing protamine sequences from a variety of mammals by a cladistic method similarly grouped horses more closely with humans than with artiodactyls (unfortunately, members of the Carnivora, such as dogs, were not included in the analyses; Ishiguro et al. 1989). This finding differs from most common interpretations of eutherian phylogeny but parallels our results for *IgM*. However, additional analyses of horse immune cell genes and proteins are needed to reach substantial conclusions about the relationship between the immune systems of horses, dogs, and humans.

In summary, we characterized seven horse *Igh-V* genes and identified a single *IgM* constant segment. Structurally, the horse genes had many similarities to *Ig* genes of other species but also differed in some ways including the presence of unreported sequences within canonical areas of the CDR. Most significantly, however, the horse *Igh-V* sequences contained a fairly conserved region within FW3 that allowed PCR amplification of different families with a single degenerate oligonucleotide. This data along with sequence information for the joining and constant region segments should provide the basis for in depth analyses of equine immune-related diseases.

Acknowledgments This work was supported by the Equine Research Laboratory with funds provided by the Oak Tree Racing Association, the State of California satellite wagering fund, and contributions from private donors.

References

Ben-Ezra, J. Variable rate of detection of immunoglobulin heavy chain V-D-J rearrangement by PCR: a systematic study of 41 B-cell non-Hodgkin's lymphomas and leukemias. *Leuk Lymphoma 7*: 289–95, 1992

Bernstein, K. E., Reddy, E. P., Alexander, C. B., and Mage, R. G. A cDNA sequence encoding a rabbit heavy chain variable region of the VHa2 allotype showing homologies with human heavy chain sequences. *Nature 300*: 74–76, 1982

Bloom, D. D., Davignon, J. L., Retter, M. W., Shlomchik, M. J., Pisetsky, D. S., Cohen, P. L., Eisenberg, R. A., and Clarke, S. H. V region gene analysis of anti-Sm hybridomas from MRL/Mp-lpr/lpr mice. *J Immunol 150*: 1591–1610, 1993

Chothia, C. and Lesk, A. M. Canonical structures for the hypervariable regions of immunoglobulins. *J Mol Biol 196:* 901–917, 1987

Chothia, C., Lesk, A. M., Gherardi, E., Tomlinson, I. M., Walter, G., Marks, J. D., Llewelyn, M. B., and Winter, G. Structural repertoire of the human VH segments. *J Mol Biol 227*: 799–817, 1992

Cook, G. P. and Tomlinson, I. M. The human immunoglobulin VH repertoire. *Immunol Today 16*: 237–242, 1995

Creighton, T. E. The folded conformations of globular proteins. *In: Proteins: Structures and Molecular Properties*, pp. 234–235, W. H. Freeman and company, NY, 1984

Felsenstein, J. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution 39*: 783–791, 1985

Harindranath, N., Donadel, G., Sigounas, G., and Notkins, A. L. Comparison of complete nucleotide sequence of the human IgM heavy chain constant region of polyreactive and monoreactive antibodies *Mol Immunol 30*: 111–112, 1993

Hein, W. R., Dudler, L., Marcuz, A., and Grossberger, D. Molecular cloning of sheep T cell receptor gamma and delta chain constant regions: unusual primary structure of gamma chain hinge segments. *Eur J Immunol 20*: 1795–1804, 1990

Honjo, T. Immunoglobulin genes. *Annu Rev Immunol 1*: 499–528, 1983

Horkvik, I., Moie, A. M., Glette, J., Male, R., Endresen, C. Cloning and sequence analysis of two isotypic IgM heavy chain genes from Atlantic salmon, *Salmo salar. L. Eur J Immunol 22*: 2957–2962, 1992

Ishiguro, H., Ichihara, Y., Namikawa, T., Nagatsu, T., Kurosawa, Y. Nucleotide sequence of *Suncus marinus* immunoglobulin μ gene and comparison with mouse and human μ genes. *FEBS Letts 247*: 317–322, 1989

Kabat, E. A. *Sequences of Proteins of Immunological Interest, 4th edn*, p. 266, US Dept. Health and Human Services, Bethesda, MD, 1987

Kirkham, P. M., Mortari, F., Newton, J. A., and Schroeder, H. W. Immunoglobulin VH clan and family predicts variable domain structure and may influence antigen binding. *EMBO J 11*: 603–609, 1992

Kuby, J. Immunoglobulins: Structure and Function. *In: Immunology, Second Edition*, pp. 109–131, W. H. Freeman and Company, NY, 1994

McCumber, L. J. and Capra, J. D. The complete amino-acid sequence of a canine mu chain. *Mol Immunol 16*: 565–570, 1978

McGuire, K. L., Duncan, W. R., and Tucker, P. W. Phylogenetic conservation of immunoglobulin heavy chains: direct comparison of hamster and mouse C mu genes *Nucleic Acids Res 13*: 5611–5628, 1985

Meek, K., Rathbun, G., Reininger, L., Jaton, J. C., Kofler, R., Tucker, P. W., and Capra, J. D. Organization of the murine immunoglobulin VH complex: placement of two new VH families (VH10 and VH11) and analysis of VH family clustering and interdigitation. *Mol Immunol 27*: 1073–1081, 1990

Navarro, P., Barbis, D. P., Antczak., D. and Butler, J. E. The complete cDNA and deduced amino acid sequence of equine IgE. *Mol Immunol 32*: 1–8, 1995

Novacek, M. J. Mammalian phylogeny: shaking the tree. *Nature 356*: 121–125, 1992

Parker, K. E., Bugeon, L., Cuturi, M.-C. and Soulillou, J.-P. Cloning of cDNA coding for the rat μ heavy chain constant region: differences between rat allotypes. *Immunogenetics 39*: 159, 1994

Patri, S. and Nau, F. Isolation and sequence of a cDNA coding for the immunoglobulin mu chain of the sheep. *Mol Immunol 29*: 829–836, 1992

Sanger, F., Nicklen, S., and Coulson, A. R. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci USA 74:* 5463–5467, 1977

Saitou, N. and Nei, M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol 4:* 406–425, 1987

Schrenzel, M. D., Watson, J. L., and Ferrick, D. A. Characterization of horse (*Equus caballus*) T-cell receptor beta chain genes. *Immunogenetics 40*: 135–144, 1994

Schrenzel, M. D. and Ferrick D. A. Horse (*Equus caballus*) T-cell receptor alpha, gamma, and delta chain genes: nucleotide sequences and tissue-specific gene expression. *Immunogenetics 42*: 112–122, 1995

Swofford, D. L. PAUP: Phylogenetic Analysis using Parsimony. 3.1 Illinois Natural History Survey, Champaign, Illinois, 1993

Takeuchi, N., Ishiguro, N., and Shinagawa, M. Molecular cloning and sequence analysis of bovine T-cell receptor gamma and delta chain genes. *Immunogenetics 35*: 89–96, 1992

Thompson, C. B. Creation of immunoglobulin diversity by intrachromasomal gene conversion. *Trends Genet 8*: 416–421, 1992

Tonegawa, S. 1983 Somatic generation of antibody diversity. *Nature 302*: 575–581, 1983

Tutter, A. and Riblet, R. Conservation of an immunoglobulin variable-region gene family indicates a specific, noncoding function. *P N A S 86:* 7460–7464, 1989

Vasquez, M., Mizuki, N., Flajnik, M. F., McKinney, E. C., and Kasahara, M. Nucleotide sequence of a nurse shark immunoglobulin heavy chain cDNA clone. *Mol Immunol 29:* 1157–1158, 1992

Wan, J. H., Trainor, K. J., Brisco, M. J., and Morley, A. A. Monoclonality in B cell lymphoma detected in paraffin wax embedded sections using the polymerase chain reaction. *J Clin Path 43:* 888–890, 1990

Annex III

GenBank: AY158087.1

Bos taurus immunoglobulin heavy chain (IGH) gene, partial sequence

- Features
- Sequence

```
LOCUS       AY158087                8259 bp    DNA     linear   MAM 10-SEP-
2003
DEFINITION  Bos taurus immunoglobulin heavy chain (IGH) gene, partial
sequence.
ACCESSION   AY158087
VERSION     AY158087.1  GI:34558505
KEYWORDS    .
SOURCE      Bos taurus (cattle)
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata;
Euteleostomi;
            Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla;
Ruminantia;
            Pecora; Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 8259)
  AUTHORS   Zhao,Y., Kacskovics,I., Pan,Q., Liberles,D.A., Geli,J.,
Davis,S.K.,
            Rabbani,H. and Hammarstrom,L.
  TITLE     Artiodactyl IgD: the missing link
  JOURNAL   J. Immunol. 169 (8), 4408-4416 (2002)
   PUBMED   12370374
REFERENCE   2  (bases 1 to 8259)
  AUTHORS   Zhao,Y., Kacskovics,I., Rabbani,H. and Hammarstrom,L.
  TITLE     Physical mapping of the bovine immunoglobulin heavy chain
constant
            region gene locus
  JOURNAL   J. Biol. Chem. 278 (37), 35024-35032 (2003)
   PUBMED   12829708
REFERENCE   3  (bases 1 to 8259)
  AUTHORS   Zhao,Y. and Hammarstrom,L.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-OCT-2002) Center for Biotechnology at Novum,
            Karolinska Institute, Huddinge, Stockholm S-141 57, Sweden
FEATURES             Location/Qualifiers
     source          1..8259
                     /organism="Bos taurus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9913"
                     /chromosome="21"
                     /map="21q23-q24"
     gene            <1..>8259
                     /gene="IGH"
                     /note="immunoglobulin heavy chain"
```

```
     misc signal     210..260
                     /gene="IGH"
                     /note="JH-PS1"
     misc signal     390..440
                     /gene="IGH"
                     /note="JH-PS2"
     misc signal     737..784
                     /gene="IGH"
                     /note="JH-PS3"
     J segment       1018..1062
                     /gene="IGH"
                     /note="JH1"
     misc signal     1401..1448
                     /gene="IGH"
                     /note="JH-PS4"
     J segment       1925..1975
                     /gene="IGH"
                     /note="JH2"
     misc feature    2849..2883
                     /gene="IGH"
                     /note="enhancer core region"
     misc feature    4360..7380
                     /gene="IGH"
                     /note="mu switch region"
     C region        7821..8141
                     /gene="IGH"
                     /note="mu CH1"
ORIGIN
        1 gggaggggca ggagcaatgt gacctcatgg gaacaaagca gatcaggaga cagcagacac
       61 tcaaatccca gaccaggggc ctgaggcggg caggtgcttg tggggtgggg ggcaagtgtc
      121 ttcaggaaga cagacttgca gctcctgggg gctccctggg gaccctgggt ggtttccgtg
      181 ctcccagcac ggggccagcc cactgtgact atgctgactt ccatctctgt ggccaggctg
      241 ccctgggcac cgtctcctca gctgaatctg gtgtctagtg tccaagtcag gggacagtgg
      301 gacttgggtg gacttgggat gtgagtcagg gccagcaagg aggtagggtg tgtttttgtg
      361 gaaaagaagc aacagaatgg aagccatgct gctgggacat ggatctctgg ggccagcgca
      421 ccccggtcac cgtgtccttg gggagtgtct cccttgcagc actgcctggg cctcaggcag
      481 ttccctgggg tccgtgcagc cccaccgtgc cctgagcaag gtagggtcct ggtccaacgg
      541 ggtcccgccc ttggatctgc cggagcccct ggagatgagc cagctgaggc ctgaggggag
      601 gacaggagct ggctgggcgg gaggaaacgt gggcagagcc cgacggggca gggtctgttc
      661 agtgagggca ggggagaccg ggcctggcct ccctagggtt catgtcttgg gggcagccag
      721 gactgtgtcc ctcagcaatg cttttgactc ctggggccag cgcgccccgg tctccatctc
      781 ctcaggtgag acggctctct gcccgctccg tcctgggccg gggaagatgt ctccaggggc
      841 cccttggtct ggggcagacc ctccgtggtc cctgggggat ctgcgtctga ggccgtttgc
      901 ctctctgccc tgttgaggtg gcgcctctgc ctgtggaaca ctggctttgc atgggggcag
      961 aggacgccag gccttgggtt tttgcacagc ccctagcggg gcccacggca ctgtgactac
     1021 gtcgatgcct ggggccaagg actcctggtc accgtctcct caggtgagtc ctcaacagcc
     1081 ctctctcctc actctctctc agggttttgg tgcactttgg ggaaaatcga gggtgtcggg
     1141 tctaaggggc ctgggcagcc cgggggtctg aaacactgag ggcccaaggg caggcttaca
     1201 gcaccgagga gcagaggctc caggcaccct ctcctcctgg gctctgtagc cagggctttc
     1261 tctgggggcc tcagcaccct tgggctctgg actcccaagg tcccagctgt gctgcctcat
     1321 gaggcacagg tgaggtggtc cggagcctca gccggccacg agctcttgcc tggggtccca
     1381 gcatcattgt caccgtgtaa caactggctc aagcactggg gtcgggaagc ctgggcactg
     1441 tctgctcagc tgagccctca ccaccccact tcactgcacc cggggagacc cagggtgtca
     1501 gagatccagg ggcattctgg aggtcaagaa agggagctgg ggagagggtt ctggtgacag
     1561 gcagagcaag acctcctcgc cccaaggaca ccgcgatgtg ggtacgaggt ggctcctttg
     1621 gcgggtctgg ctgcctgact tgagcaggac caggggcttc cgtcgctttc tggggcaggc
     1681 ggctgctcga ggctggactt aggagtctgt ggttcacggt cagccggccc aggcaggcag
     1741 tggtctgcct ctggggacca gaatgggaca tagtgtctct ggcagagtca gggtcacaca
     1801 caagtgactt tgaccgacgg cttccctgtg gcgcctggag atggcgtggg ggcccaggcg
     1861 cctcgagcct tgccaggctc ccgaggtttt tgttgggcga ggctggagat atcaccactg
     1921 tgattactat agtatatatg tttgcggccg agggatcgag gtcaccgtct cctcaggtaa
```

1981 gagcggccca tccagggcct ttgctttatc tcatttcgtg cgatttttct gagcatcact
2041 gtccggtcct ctgatgtgtc cttgtcccct cctccccggg gggactcggc agactggcca
2101 ggaggggacc agctgcccta tgcatttcag agtctctatt ttctgatgcc tttaaaaaat
2161 cagaatttca ttggcattca gaggggggctt gggcgggaag ggccaccagt gggggaggcc
2221 caggcccccc ttggcagcag ggcagcttgg gatgcggtgg aaggcaactt gccacggtcc
2281 tagcatctgc agaggagcat gtctggataa ttagggcctc aggaacgctg cccgcggtgg
2341 gcacagagaa cgccctcctc gggtgaggtg tgttctgcac tagactgtgt ttaaaattct
2401 ttattgggta ggaagagaat tgtctaggtg aggacggaca tgcagtgtcc cgatcgtggc
2461 gagagagggg agcccgggga ggtgacgggc gctgggcttt gtgaggccag tctaagagag
2521 aaaggccgtt cgccagagga ggtgtgcttg cgaacaccaa gacagggcat ctctgaagcg
2581 attcctgata gtctgaaaaa ttgaaacttt aaaaagaaat gtttaaagta ttttaaattt
2641 ttatcattta attaacaacc gcaaatcatg gctttggaga gttgagacag gtacaagttt
2701 ggccgaaaag tactaactag gttccatcgg ccctcggccc caattcaggg ctgttttgag
2761 aataataaat tcagcttatt tttttaatgt aattggtggt gccgagttag tcaagatggc
2821 cacgggcagg actgaccacc tgcagcaggt ggcaggaagc tgagagtctg tttttggaag
2881 caagaaaaca cagttggtaa atttatagct tctggttccc aaaaggtggt ttgcggctgg
2941 ttttgcccag ccccacagaa ccgaaagtgt tcccctgagc ggagcaacac ctggctaatt
3001 tgcatttcta aaacaaggcg cagatgctga ccgaaactgg aaggttcctc ttttaactat
3061 ttgaatttac ttcagcttta gcttatcaac tgctcactta tattcatttt caaagtcgat
3121 gtttaagaaa ttaatttgtc tcaggtgtat tttatcaatg caatcgtatg ctctttaatg
3181 agttactcct tcaaggtagt taagctgaga ctgtgcgcac gccgtgtgca ctgagtgtgc
3241 tcacaatcgt agctggccgg ctccctcctg gcgatgaagc agctttctca gtgcggctcc
3301 ctcactgcag aatgttcagg gccgcctttc gggtacccat ctcaggacag acttgggtga
3361 aacgggggca ttcgtcgtga gttactctgc tacactacta gctgggggat tgagaggatg
3421 cccagtgggg tgggagtcag cggggcctgg gacggcccgc agctggggac gtgcgggcct
3481 ggcctctgcc gccaggccag tgggcaggtc acgagttagc acagtgaagg gattcagagg
3541 aatccaccca ggtcagagag cccacgggac atgccaggga ctaggcgccg agctgtgagg
3601 ggcaggagcc ggtgcagcca gactcgctgc cccgggacgt ttgtatgagt tgtttttgtt
3661 tctacttgaa agggaatctg tctttaccta ctgcctgcga tgccggggtg cttcacaccc
3721 acagggacca cttcctgcgt ccacacttag gtgttcaatg cagctttggg gttcagatgg
3781 tctctgtttg ctagagagta tcctctttct cagcagaact ttgtgataaa gacagaattg
3841 ggagcaggtt cagataagat gcatttcccg tggtttttaa tgaggacatt ggccagagag
3901 cagttagtca ccaggactct tttagactcg gggcaggatc ttaggcccac ggcatgtgcc
3961 caggaggcca ggccccaggg acgtatgccc aggctctcct ctcccagaga agcagcgctg
4021 cctgctcagc agagacaggg cctctgtctg agggtggcaa gcgtgtctca gcctgtgaga
4081 tgcagtaggg aagctaaagt cgtcacgatg cttttcaccc agagtggagg gtggtggtga
4141 ctgacggctg ggagaagggc cggctgggcc tggtgtgggg atggaggcat cgggcaccca
4201 ggagaagcag gaagaggctg gaaggggtga gcccagctgg gcagcatccg cagaccaggg
4261 gctttagccg aggccacggg tccgggatct gacccttttct cgtcaggacc actgggcttt
4321 cacgtctgtg agggcagcct gggccaccga cagcaggtga gctggtctgg gctgaactta
4381 gctgagccag gttgggctgg gctgagctgc actgagctgc gctgggctga gctgggttga
4441 gctgggctga gcagagctgg gcagagtggg tgagctgggc tgagctgagc taagctgaac
4501 tggagtgggt gagctgggct gggctgggct gggctgggct gggttgagct gggctgggct
4561 gagctgcact gagctgggct aagctgggtt gggctgggct tagtggggct gggcggagtg
4621 ggtgagctgg gctgagctgg attaggctga gttgagcgga gctgggcaga gtgggtgagc
4681 tgggctgagc tgggttaggc tgagttgagc tgggctcagc tgggctgagc tgggtggagt
4741 gggtgagctg ggctgaactg ggttaggctg agttgagcta ggctcagctg ggctgcgctg
4801 agctgcactg agctgagctg ggctgggctg ggctgagcgg agctgggagg agtggatgag
4861 ctgagctaag ctgggttagg ctgagttgag ctgggcttag ctgggctggg ctgagcggag
4921 ctgggaggag tgggtgagct gagctaagct gggttaggct gagttgagct ggactcagct
4981 gggctgcgct tagctgggct gggctgagca gggctgagca gagctgggcg gagtgggtga
5041 gctgggctga gctgggttag gctgggctga gctgggttag gctgcgatga gctgggctga
5101 acggagctgg gcggagtggg tgagctgagc tgggctgggc tgagctgggc agagtgggtg
5161 agctgggctg agctgggttg ggctgagttg agataagctc agctcggctg agctgggcag
5221 agtgggtgag ctggactgag ctgagctggg ctgggctgag ctgggcggag cagagctggg
5281 cagagtgggt gagctgggct acactgggct gggctgagct gcactgagct gggctgggct
5341 gagctgggct gagcggagct gggaggagtg ggtgagctgg gctaagctgg gttaggctga
5401 gttgagctgg actcagctgg gctgcgctta gctgggctgg gctgagcaga gctgggcgga
5461 gtgggtgagc tgggctgagc tgggttaggc tgagctgggc tgagcggagc tgggcagagt
5521 gggtgagctg ggttgggctg ggctgcgctg agctgggctg agcggagctg ggcagagtgg
5581 gtgagctggg ctgagctggg ttaggctgag ctgagcggag ctgggcggag tgggtgagct

```
5641 gggttaggct gggctgggct gggctgggct gggcagagtg ggtgagctgg actgagctga
5701 gctgggctgg gctgacctgg gctgagcaga gctgggcgga gtgggtgagc tgggctgggt
5761 tgggctgggc tgagctgagc taagctgaac tgggctgggt tgagctgggt ggagtgggtg
5821 agctgggctg agctgggttg agctgggctg agctgaatgg agtggctgag ctgggttgag
5881 ctgggcagag tgggtgagct gggctgagct aggtggagtg ggtgagctga gctgagctga
5941 gctgaatgga gtggctgagc tgagctgggt tgagctgggc agagtgggtg agctgggctg
6001 agctgggtgg ggtgggtgag ctgagctgag ctgagctgaa tggagtggct gagctgagct
6061 gggttgagct gggcagagtg ggtgagctgg gctgagctag gtggagtggg tgagctgagc
6121 tgagctgagc tgaatggagt ggctgagctg agctgggttg agctgggcag agtggctgag
6181 ctaggctgag ctgggttgag ctgggctgag ctgaatggag tagctgagct gggctgaggg
6241 gggctgagct gggccaagtg tgtgggctgg gctgggctgg gcttagttgg gttgacctgg
6301 gctgagctga gctgggcttg gttgagctgg gctgagcttg gctgagcttt gctgggctga
6361 gctgggcaga gtgggtgagc tgggccgggt tgggttgagc tggcctgagc tgggttgagc
6421 tggcctgagc tgggctgagt gggtgagctg ggctgggctg ggctgggctg ggctgagttg
6481 gtgagctggg ctgggctggg ctgagctggg gtgagctggg ctgggctggg ttgagctggc
6541 ctgagctggg cagagtgggt gagctgggct gaactgagct gagctgggtt gagcaaggct
6601 gtgttgagct gggcagagtg ggtgagctag gctgagctgg gctgagctgg actgagctcg
6661 gctgaactga gctgagttga gttgggctga gctggcctgg gctgggctga gctgggctaa
6721 actgagctga gctgggttga gctgggctgg gttgagctgg gctgggttga gctgggctgg
6781 gttgagctgg gcagagtggg tgagctaggc tgggctgggc tgtgctggca gagtgggtga
6841 gctgggctga gctgggttga gctgaggtga gctgggctga actgagctgc actgtgttga
6901 gctgcgcaga ttgggtgagc tgggctgaac tgagctgagc tgggctgttt aaactgggca
6961 gagtgggtga gctaggctga gctgggctga actgagctga gttgagttgg gctgagctgg
7021 gctgaactga gctgacccag gttgagctga gctgggttga gctgggcagg gctgggctgt
7081 gctggcagat tgggtgagct aggctgagct gagtgggctg agctgggctg tttaagctgg
7141 gcagagtggg tgagctaggc tgagctgggc tgaactgagc tgagctgggc tgagctgggc
7201 ggagtgggtg agctgggctg agctgagctg ggcagaactg ggcttagctg gctgagcagg
7261 tgagctgggc tgagagggtt gagctgagct gagctgaact gagctgggtt gagctggcct
7321 ggattgggct gatctggatt gaactgggct gaactgagct gagctgggct gagctgggct
7381 gggctagcct gggcttaact ggggtaagct gggttaacca tggtgggctg agactgggtg
7441 aacagaaggg aacccactga gctagcttga cttgggcaaa ttgttgagat gctttgggta
7501 catttaacct ccactgaggc tggtgggtct agatggactg agctgacccc actggactgt
7561 tctggtcaag agtgggctgg cctcagtgat ctgattgtgc tgaggactgg gagcctgagt
7621 gtgtatcagt cttattgatg accccagacc cccagggcag gccccaggag gctaatcgcg
7681 ggcagtgagg gtgggctggc agggctgagt gtgccctccg ctccactgtc ccagagagaa
7741 ggtggagctg cccacacccc cgaccagcag gatgcctctc acccccctct tctcctgtgt
7801 cctctctcgg gtccccagaa ggtgaatcac acccgagagt cttccccctg gtgtcctgcg
7861 tgagctcgcc atccgatgag agcacggtgg ccctgggctg cctggcccgg gacttcgtgc
7921 ccaattcagt cagcttctcc tggaagttca acaacagcac agtcagcagc gagagattct
7981 ggaccttccc cgaagtcctg agggacggct tgtggtcggc ctcctctcag gtggtcctgc
8041 cctcctcaag cgcctttcaa gggccggatg actacctggt gtgcgaagtc cagcacccca
8101 agggaggaaa gaccgtcggc accgtgaggg tgatcgctac aagtgagtca ggtacgtccc
8161 gtggttgggt gcgggggggag ggtccaggcc ccgctaaccc cttgtccttc tctgcagagg
8221 cggaagtgct gtccccagtc gtgagtgtct ttgtcccgc
```

Annex IV

GenBank: M13800.1

Rat Ig germline mu H-chain C-region gene, exon 2

- Comment
- Features
- Sequence

```
LOCUS       RATIGCF3                659 bp    DNA     linear   ROD 27-APR-
1993
DEFINITION  Rat Ig germline mu H-chain C-region gene, exon 2.
ACCESSION   M13800
VERSION     M13800.1  GI:204709
KEYWORDS    C-region; germline; immunoglobulin heavy chain; immunoglobulin
            mu-chain.
SEGMENT     3 of 3
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata;
Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 659)
  AUTHORS   Bruggemann,M., Free,J., Diamond,A., Howard,J., Cobbold,S. and
            Waldmann,H.
  TITLE     Immunoglobulin heavy chain locus of the rat: striking homology
to
            mouse antibody genes
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 83 (16), 6075-6079 (1986)
   PUBMED   3016742
COMMENT     Original source text: Rat (PVG) liver DNA, clones 22-3-1, 5-2-
1,
            45-3-1, 31-1-1, 18-1-3, and 24-3-1.
            Computer-readable copy of the sequence in [1] was kindly
provided
            by M.Brueggemann, 19-NOV-1986.
FEATURES             Location/Qualifiers
     source          1..659
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
     gene            1..659
                     /gene="IgM"
     intron          <1..77
                     /gene="IgM"
                     /note="J1-C intron"
                     /number=1
     CDS             78..416
                     /gene="IgM"
                     /note="C-region"
```

```
                     /number=2
                     /codon_start=3
                     /product="immunoglobulin mu-chain"
                     /protein_id="AAA41372.1"
                     /db_xref="GI:554448"

/translation="VVEMNPNVSVFIPPRDAFSGPAPRKSRLICEATNFSPKQITVSW

LQDGKPVKSGFTTEPVTVEAKGSRPQTYKVISTLTITESDWLNLNVFTCRVDHRGLTC
                     LKNVSSTCAA"
     intron          417..>659
                     /gene="IgM"
                     /number=2
ORIGIN      275 bp upstream of BamHI site; about 7 kb after segment 2.
        1 ggcccaggca tggcccagag ggagcagcga gtgggtctta agccagcctg agctcacacc
       61 tcaacctttc attccagctg tcgttgagat gaaccccaat gtgagtgtgt tcattccacc
      121 acgtgatgcc ttctctggcc ctgcaccccg caagtccaga ctcatctgcg aggccaccaa
      181 cttcagtccc aaacagatca cagtatcctg gctacaggat gggaagcctg tgaaatctgg
      241 cttcaccaca gagccagtga ctgtcgaggc caaaggatcc agaccccaaa cctacaaggt
      301 cataagcaca ctgaccatca ctgaaagcga ctggctgaac ctgaatgtgt tcacctgccg
      361 cgtggatcac aggggtctca cctgcttgaa gaacgtgtcc tccacatgcg ctgccagtga
      421 gtagcctgtg ctaagcccaa tgcctagccc tcccacatta gagcagtcct cctacggttg
      481 tggccaatgc cacccagaca tggtcatttg cttcttgagc cttggcttcc aacagtggcc
      541 aaggccaagg atgagcagta ggcagcaggg ggatgagagt cagatggagg gaatcagcat
      601 cttcccttaa gcagatttgg aagatggaga ctgagctttt atccaacttc acaactaga
```

Annex V

GenBank: X02804.1

Syrian hamster Cmu-gene for immunoglobulin heavy chain

- Comment
- Features
- Sequence

```
LOCUS       X02804                  2641 bp    DNA     linear   ROD 14-NOV-
2006
DEFINITION  Syrian hamster Cmu-gene for immunoglobulin heavy chain.
ACCESSION   X02804 M23867
VERSION     X02804.1  GI:49635
KEYWORDS    constant region; Ig heavy chain; immunoglobulin.
SOURCE      Mesocricetus auratus (golden hamster)
  ORGANISM  Mesocricetus auratus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata;
Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Cricetidae; Cricetinae; Mesocricetus.
REFERENCE   1  (bases 1 to 2641)
  AUTHORS   McGuire,K.L., Duncan,W.R. and Tucker,P.W.
  TITLE     Phylogenetic conservation of immunoglobulin heavy chains:
direct
            comparison of hamster and mouse Cmu genes
  JOURNAL   Nucleic Acids Res. 13 (15), 5611-5628 (1985)
   PUBMED   2994005
COMMENT     On Feb 24, 2005 this sequence version replaced gi:191395.
            Data kindly reviewed (19-FEB-1986) by P.W. Tucker.
FEATURES             Location/Qualifiers
     source          1..2641
                     /organism="Mesocricetus auratus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10036"
     gene            join(<466..778,882..1220,1474..1791,1906..2300)
                     /gene="Cmu"
     CDS             join(<466..778,882..1220,1474..1791,1906..2300)
                     /gene="Cmu"
                     /codon_start=1
                     /product="immunoglobulin heavy chain"
                     /protein_id="CAA26574.1"
                     /db_xref="GI:1333850"
                     /db_xref="GOA:P06337"
                     /db_xref="InterPro:IPR003006"
                     /db_xref="InterPro:IPR003597"
                     /db_xref="InterPro:IPR007110"
                     /db_xref="InterPro:IPR013783"
                     /db_xref="UniProtKB/Swiss-Prot:P06337"

/translation="SPSSPTVFPLVSCESPLSDENLVAMGCLARDFLPSSISFSWNYQ

NKSEVNQGVRTFPTLRMGEKYAATSQVFLPPKSVLEGSDEYLVCKVHHGNTNKDLRVP
```

IPVVTEMNPNVSVFVPSRDAFSGPAPRKSRLFCEASNFSPKQITVSWLRDGKPVKSGF

TTEPVTPEDRGSGPRTYKVISTLTITESDWLNLSVYTCRVDHRGLTFWKNVSSTCAAS

PSTDIQAFPIPPSFVGIFLNKSATLTCLVTNLATYDTLNISWSSRSGEPLETKTKLTE

SHPNGTFSAIGEANVCVEDWDSGKEFVCTVTHRDLPSPQKKFISKPREMNKTPPAVYQ

QPLAREQLILRESATVTCLVKGFSPADIFVQWLQRGQPLSQDKYVTSAPMREPQAPHL

YFTHSVLTVTEEEWNSGETYTCVVGHEALPHMVTERTVDRSTGKPTLYNVSLIMSDAG
GTCY"

exon	466..778
	/gene="Cmu"
	/number=1
misc feature	598..606
	/gene="Cmu"
	/note="pot. glycosylation site"
intron	779..881
	/gene="Cmu"
	/number=1
exon	882..1220
	/gene="Cmu"
	/number=2
misc feature	902..910
	/gene="Cmu"
	/note="pot. glycosylation site"
misc feature	983..991
	/gene="Cmu"
	/note="pot. glycosylation site"
misc feature	1142..1150
	/gene="Cmu"
	/note="pot. glycosylation site"
misc feature	1196..1204
	/gene="Cmu"
	/note="pot. glycosylation site"
intron	1221..1473
	/gene="Cmu"
	/number=2
exon	1474..1791
	/gene="Cmu"
	/number=3
misc feature	1533..1541
	/gene="Cmu"
	/note="pot. glycosylation site"
misc feature	1590..1598
	/gene="Cmu"
	/note="pot. glycosylation site"
misc feature	1659..1667
	/gene="Cmu"
	/note="pot. glycosylation site"
intron	1792..1905
	/gene="Cmu"
	/number=3
exon	1906..2421
	/number=4
misc feature	1911..1919
	/gene="Cmu"
	/note="pot. glycosylation site"
misc feature	2256..2264

```
                     /gene="Cmu"
                     /note="pot. glycosylation site"
     polyA signal    2401..2406
     polyA site      2421
ORIGIN
        1 cttgagcccc tcccccacca ccaccacctc ccgatgggct tctattgttt tggaatagga
       61 atgcttccct cagctgactt ctcagaggga agcctagcag gacagggcta agcctatttg
      121 agaagcgggg atcctggact gtccctgaga acaggtgtca gtgccccgcc cctggagctg
      181 gctgacgatg gcggaagggc caggtccact tagtctagca ggcatctgtc cctgtccatg
      241 aaggcttaca gtgttctatg gctggaaaag atactccgta agcctgcagc aacttcgtcc
      301 tgtaggccag gttcgcctac caacatatct agccctgtct gccttaagag caacagccac
      361 agaggcgggg gggaatgcag atagcccctg actggtagtg gacaagcaga cagattctta
      421 ctcctcctcc tccgtttctc ttttatccct ctctggtcct cagagagtcc gtcctcccca
      481 actgtcttcc ccctggtctc ctgcgagagc cccctgtctg atgagaattt ggtggccatg
      541 ggctgcctgg cacgggactt cctgcccagc tccatttcct tctcctggaa ctaccagaac
      601 aagtctgaag tcaaccaggg tgtcagaacc ttcccaacac tgaggatggg agagaagtat
      661 gcagccacct ctcaggtgtt cctgcctccc aagagtgtcc tcgagggttc agatgaatac
      721 ctggtgtgca aagtgcacca tggcaacacc aacaaagatc tacgcgtgcc cattccaggt
      781 aagaaccagc cagtgggggc agaggcccag gcatggccca gaggaggccg ggggctgcag
      841 cttagagcaa tctgagctca caccttgtcc tttcactcca gtcgtcacag agatgaaccc
      901 caacgtgagt gtgttcgttc catcccggga cgccttctct ggccctgccc cacgaaagtc
      961 cagactcttc tgcgaagcct ctaacttcag tcccaaacaa atcacagtgt cctggctacg
     1021 tgacgggaag cctgtgaagt ctggcttcac cacagagcca gtgactcctg aagacagagg
     1081 ttccggaccc cgaacctaca aggtcataag cacactcacg atcaccgaaa gcgactggct
     1141 gaacctgagt gtgtacacct gccgtgtgga tcacaggggt ctcaccttct ggaagaatgt
     1201 gtcctccaca tgtgctgcca gtgagtgccc tgtcctagcc cagtggctat ccctcccaga
     1261 tcagggccac cctcttacaa gtaaggctaa tgccaccctg acatggccat ttgctccttg
     1321 agccttggtt tcaagggcca agaccagtgt cttcccttaa gcagaacctc ggaggttcag
     1381 taaatcagtt ggtactattt ggggatagag acttaagctt ttctccgact tcacaacttg
     1441 acagccactg acacaatgtt ttcttgactg caggcccatc cacagacatc caagccttcc
     1501 ctatcccccc ctcctttgtt ggcatcttcc tcaacaagtc tgccacgctg acctgcctgg
     1561 tcacaaacct ggcaacctac gacaccctga acatctcctg gtcttcccga agtggtgagc
     1621 cactggaaac caaaactaaa ctcactgaaa gccaccccaa tggcaccttc agtgccatag
     1681 gtgaggctaa tgtttgtgtg gaagactggg atagcgggaa ggagttcgtg tgcaccgtga
     1741 cccacaggga tctgccttca ccacaaaaga aattcatctc aaagcccaga ggtatgcatt
     1801 tcccatttcc ttgtcctctg gtcagctcca gggttcttcc tctgtgtcac agggatggca
     1861 ggtccccttc tgccaccata tcctcaccac tgtctctact tgcagagatg aacaagaccc
     1921 cccctgctgt gtaccagcag ccactagccc gagaacaact gatcctgagg gagtcagcca
     1981 cagtcacctg cctggtgaag ggtttctccc ccgcagacat ctttgtgcag tggcttcaga
     2041 gggggcaacc cttgtcccaa gacaagtatg tgaccagcgc cccaatgcgg gagccccagg
     2101 ccccacacct ctacttcacc cacagcgtcc tgactgtgac agaggaggaa tggaactccg
     2161 gggagaccta cacctgtgtg gtaggccacg aggccctgcc acacatggtg accgagagga
     2221 ccgtggacag gtccaccggt aaacccacac tgtacaatgt atcgctgatc atgtctgatg
     2281 caggtggcac ctgctactga ccatgctagc ccccaaccag gcaggctctg gggggccagt
     2341 tgcactgttt acacaaacta accatgtcag agtgagatgt tgcatcttat aaaaactaga
     2401 aataaaaaaa aagatccact caaaagttgc tggttttgat tttgcaatgc ttttggctgc
     2461 cgaggcaacg gctgtgttgt gtttgccttg cacgcaccct gcagatttgc ctctgccctg
     2521 acccttcctc tgtctgcaca gttgcctcct tctgtgtgaa ctcacagagt gtaaggatgc
     2581 catccctcca gctgcttcca gatacagggc tggaagcgtg gctaccctgg tgcaggcaga
     2641 a
```